(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 November 2001 (29.11.2001)

PCT

(10) International Publication Number
WO 01/90122 A2

(51) International Patent Classification[7]: C07H

(21) International Application Number: PCT/US01/16905

(22) International Filing Date: 23 May 2001 (23.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/206,353 23 May 2000 (23.05.2000) US

(71) Applicant *(for all designated States except US)*: **GENAISSANCE PHARMACEUTICALS, INC.** [US/US]; Five Science Park, New Haven, CT 06511 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **CHEW, Anne** [US/US]; 1477 Beacon Street, #64, Brookline, MA 02446 (US). **CHOI, Julie, Y.** [US/US]; 38 Elizabeth Street, West Haven, CT 06516 (US). **MESSER, Chad** [US/US]; 41 Manor Street, Hamden, CT 06517 (US).

(74) Agents: **HENDERSON, Melodie, W.** et al.; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/90122 A2

(54) Title: HAPLOTYPES OF THE CTLA4 GENE

(57) Abstract: Novel genetic variants of the Cytotoxic T-Lymphocyte-Associated Protein 4 (CTLA4) gene are described. Various genotypes, haplotypes, and haplotype pairs that exist in the general United States population are disclosed for the CTLA4 gene. Compositions and methods for haplotyping and/or genotyping the CTLA4 gene in an individual are also disclosed. Polynucleotides defined by the sequence the haplotypes disclosed herein are also described.

HAPLOTYPES OF THE CTLA4 GENE

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/206,353 filed May 23, 2000.

FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically-important proteins. In particular, this invention provides genetic variants of the human cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying, cloning, and expressing an important target protein related to the disease. A determination of whether an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is then made. Then, vast numbers of compounds are screened against the target protein to find new potential drugs. The desired outcome of this process is a lead compound that is specific for the target, thereby reducing the incidence of the undesired side effects usually caused by activity at non-intended targets. The lead compound identified in this screening process then undergoes further *in vitro* and *in vivo* testing to determine its absorption, disposition, metabolism and toxicological profiles. Typically, this testing involves use of cell lines and animal models with limited, if any, genetic diversity.

What this approach fails to consider, however, is that natural genetic variability exists between individuals in any and every population with respect to pharmaceutically-important proteins, including the protein targets of candidate drugs, the enzymes that metabolize these drugs and the proteins whose activity is modulated by such drug targets. Subtle alteration(s) in the primary nucleotide sequence of a gene encoding a pharmaceutically-important protein may be manifested as significant variation in expression, structure and/or function of the protein. Such alterations may explain the relatively high degree of uncertainty inherent in the treatment of individuals with a drug whose design is based upon a single representative example of the target or enzyme(s) involved in metabolizing the drug. For example, it is well-established that some drugs frequently have lower efficacy in some individuals than others, which means such individuals and their physicians must weigh the possible benefit of a larger dosage against a greater risk of side effects. Also, there is significant variation in how well people metabolize drugs and other exogenous chemicals, resulting in substantial interindividual variation in the toxicity and/or efficacy of such exogenous substances (Evans et al., 1999, *Science* 286:487-491). This variability in efficacy or toxicity of a drug in genetically-diverse patients makes many drugs ineffectiv or even dangerous in certain groups of the population, leading to th failure of such drugs in clinical trials or their early withdrawal from the market even though they could be highly beneficial for

myocardial failure caused by T-cell blast infiltration (Waterhouse et al., *Science* 1995; 270:985-988). These results suggest that CTLA4 may play an inhibitory role in regulating lymphocyte expansion.

The cytotoxic T-lymphocyte-associated protein 4 gene is located on chromosome 2q33 and contains 4 exons that encode a 223 amino acid protein. Reference sequences for the CTLA4 gene (Genaissance Reference No. 743670; SEQ ID NO: 1), coding sequence (GenBank Accession No:NM_005214.1), and protein are shown in Figures 1, 2 and 3, respectively.

Several variations have been identified in the CTLA4 gene that may be related to various disorders. A polymorphism of adenine or guanine at nucleotide position 37902 in Figure 1 results in an amino acid variation of threonine or alanine at amino acid position 17 in Figure 3 (HGBASE:SNP000000387). Donner et al., (*J Clin Endocrinol. Metab* 1997; 82:4130-4132) reported that patients with Hashimoto thyroiditis had higher frequencies of the Thr17Ala mutation. However, there was no significant variation in patients with Addison's Disease and control subjects. The Thr17Ala mutation has also been shown to be associated with Grave's Disease in a dataset of white Caucasian subjects (Heward et al., *J Clin Endocrinol. Metab* 1999; 84:2398-2401). Djilali-Saiah et al. (*Gut* 1998; 43:187-189) found that in French Caucasian patients with Celiac disease, which is characterized by immunologically mediated intestinal injury following ingestion of gluten, the Thr17Ala mutation was found with greater frequency in patients than in controls. These results suggest that the location of this polymorphism on the gene is critical to the function of the CTLA4 protein.

The association between Thr17Ala polymorphism and insulin dependent diabetes mellitus (IDDM) has also been studied in numerous populations. Associations have been observed several populations including Asian, Mexican-American, and certain Caucasian populations (Awata et al., *Diabetes* 1998; 47:128-129; Donner et al., *J Clin Endocrinol. Metab* 1997; 82:143-146; Lee et al., *Clin Endocrinol. (Oxf)* 2000; 52:153-157; Marron et al., *Hum Mol. Genet* 1997; 6:1275-1282), while a lack of association has been observed in other Caucasian groups (Owerbach et al., *Diabetes* 1997; 46:1069-1074). Chistiakov et al. tested whether or not the CTLA4 gene was a susceptibility marker for IDDM in a Russian population. In the case of the codon 17 polymorphism, the alanine allele was associated with the disease. The authors also examined an $(AT)_n$ microsatellite marker in the 3' untranslated region of the gene, and examined the transmission of 17 alleles varying from 92 to 130 bp in length. The transmission of three alleles was significantly different for diabetic and non-diabetic offspring (Chistiakov et al., *supra*). Therefore, the CTLA4 gene is associated with IDDM in a Russian population.

Because of the potential for variation in the CTLA4 gene to affect the expression and function of the encoded protein, it would be useful to know whether additional polymorphisms exist in the CTLA4 gene, as well as how such polymorphisms are combined in different copies of the gene. Such information could be applied for studying the biological function of CTLA4 as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the CTLA4 genotype, haplotype, or haplotype pair. In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. In a particularly preferred embodiment, the CTLA4 haplotype is selected from the haplotypes shown in Table 4, or a sub-haplotype thereof. Such methods have applicability in developing diagnostic tests and therapeutic treatments for autoimmune disorders.

In yet another embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the CTLA4 gene or a fragment thereof. The reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of thymine at PS1, guanine at PS3, cytosine at PS4, thymine at PS5, guanine at PS6 and guanine at PS7. In a preferred embodiment, the polymorphic variant comprises an additional polymorphism of guanine at PS2.

A particularly preferred polymorphic variant is an isogene of the CTLA4 gene. A CTLA4 isogene of the invention comprises cytosine or thymine at PS1, adenine or guanine at PS2, adenine or guanine at PS3, thymine or cytosine at PS4, cytosine or thymine at PS5, adenine or guanine at PS6 and adenine or guanine at PS7. The invention also provides a collection of CTLA4 isogenes, referred to herein as a CTLA4 genome anthology.

Polynucleotides complementary to these CTLA4 genomic DNA variants are also provided by the invention. It is believed that polymorphic variants of the CTLA4 gene will be useful in studying the expression and function of CTLA4, and in expressing CTLA4 protein for use in screening for candidate drugs to treat diseases related to CTLA4 activity.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express CTLA4 for protein structure analysis and drug binding studies.

The present invention also provides nonhuman transgenic animals comprising one of the CTLA4 polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the CTLA4 isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against CTLA4 protein, and for testing the efficacy of therapeutic agents and compounds for autoimmune disorders in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the CTLA4 gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes the polymorphisms, the genotypes and the haplotypes identified for the CTLA4 gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing CTLA4 haplotypes organized according to their evolutionary relationships.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the CTLA4 gene (Genaissance Reference No. 743670; contiguous lines; SEQ ID NO:1), with the start and stop positions of each region of coding sequence indicated with a bracket ([or]) and the numerical position below the sequence and the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:36 is equivalent to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25).

Figure 2 illustrates a reference sequence for the CTLA4 coding sequence (contiguous lines; SEQ ID NO:2), with the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the CTLA4 protein (contiguous lines; SEQ ID NO:3), with the variant amino acid(s) caused by the polymorphism(s) of Figure 2 positioned below the polymorphic site in the sequence.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the CTLA4 gene. As described in more detail below, the inventors herein discovered 8 isogenes of the CTLA4 gene by characterizing the CTLA4 gene found in genomic DNAs isolated from an Index Repository that contains immortalized cell lines from one chimpanzee and 93 human individuals. The human individuals included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (21 individuals), African descent (20 individuals), Asian (20 individuals), or Hispanic/Latino (18 individuals). To the extent possible, the members of this reference population were organized into population subgroups by their self-identified ethnogeographic origin as shown in Table 1 below.

Gene - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

Genotype – An unphased 5′ to 3′ sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

Full-genotype – The unphased 5′ to 3′ sequence of nucleotide pairs found at all polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Sub-genotype – The unphased 5′ to 3′ sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Genotyping – A process for determining a genotype of an individual.

Haplotype – A 5′ to 3′ sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

Full-haplotype – The 5′ to 3′ sequence of nucleotides found at all polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Sub-haplotype – The 5′ to 3′ sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Haplotype pair – The two haplotypes found for a locus in a single individual.

Haplotyping – A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

Haplotype data - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

Isoform – A particular form of a gene, mRNA, cDNA or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

Isogene – One of the isoforms of a gene found in a population. An isogene contains all of the polymorphisms present in the particular isoform of the gene.

Isolated – As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with the methods of the present invention.

Locus - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature.

Naturally-occurring – A term used to designate that the object it is applied to, e.g., naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

Nucleotide pair – The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

Phased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

Polymorphic site (PS) – A position within a locus at which at least two alternative sequences are found in a population, the most frequent of which has a frequency of no more than 99%.

Polymorphic variant – A gene, mRNA, cDNA, polypeptide or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

Polymorphism – The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

Polymorphism data – Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

Polymorphism Database – A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

Polynucleotide – A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

Population Group – A group of individuals sharing a common ethnogeographic origin.

Reference Population – A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

Single Nucleotide Polymorphism (SNP) – Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

Subject – A human individual whose genotypes or haplotypes or response to treatment or disease state are to be determined.

Treatment - A stimulus administered internally or externally to a subject.

Unphased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

As discussed above, information on the identity of genotypes and haplotypes for the CTLA4 gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is expected to be useful for a variety of drug discovery and development applications. Thus, the invention also provides compositions and methods for detecting the novel CTLA4 polymorphisms and haplotypes identified herein.

The compositions comprise at least one CTLA4 genotyping oligonucleotide. In one embodiment, a CTLA4 genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that is located close to, or that contains, one of the novel polymorphic sites described herein. As used herein, the term "oligonucleotide" refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of a CTLA4 polynucleotide, i.e., a CTLA4 isogene. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with a non-target region or a non-CTLA4 polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the CTLA4 gene using the polymorphism information provided herein in conjunction with the known sequence information for the CTLA4 gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or "complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency

conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5′ end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the oligonucleotide probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as temperatures for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in PCR Protocols, A Guide to Methods and Applications, Academic Press, 1990 and Ruaño et al., 87 *Proc. Natl. Acad. Sci. USA* 6296-6300, 1990. Typically, an ASO will be perfectly complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotides of the invention include ASO probes and ASO primers. ASO probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15mer, the 8th or 9th position in a 16mer, and the 10th or 11th position in a 20mer). An ASO primer of the invention has a 3′ terminal nucleotide, or preferably a 3′ penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. ASO probes and primers hybridizing to either the coding or noncoding strand are contemplated by the invention.

ASO probes and primers listed below use the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25) at the position of the polymorphic site to represent the two alternative allelic variants observed at that polymorphic site.

A preferred ASO probe for detecting CTLA4 gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

AGATCCTYAAAGTGA (SEQ ID NO:4) and its complement,
CAGTCAARGGCAGTG (SEQ ID NO:5) and its complement,
CACTGAGYTGACACC (SEQ ID NO:6) and its complement,
CTAGAACYGTAGGCA (SEQ ID NO:7) and its complement,

TTTTAATRGCTGAAT (SEQ ID NO:8) and its complement, and
GCTGTGARCATTCAT (SEQ ID NO:9) and its complement.

A preferred ASO primer for detecting CTLA4 gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

TTATCCAGATCCTYA (SEQ ID NO:10); TCATGTTCACTTTRA (SEQ ID NO:11);
TTTCAGCAGTCAARG (SEQ ID NO:12); ATAAATCACTGCCYT (SEQ ID NO:13);
CCATTTCACTGAGYT (SEQ ID NO:14); GCAACAGGTGTCARC (SEQ ID NO:15);
AACGCACTAGAACYG (SEQ ID NO:16); TGCCAATGCCTACRG (SEQ ID NO:17);
AATAAATTTTAATRG (SEQ ID NO:18); TTCTTGATTCAGCYA (SEQ ID NO:19);
CTGTATGCTGTGARC (SEQ ID NO:20); and TTAAAAATGAATGYT (SEQ ID NO:21).

Other genotyping oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms described herein and therefore such genotyping oligonucleotides are referred to herein as "primer-extension oligonucleotides". In a preferred embodiment, the 3′-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site.

A particularly preferred oligonucleotide primer for detecting CTLA4 gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

TCCAGATCCT (SEQ ID NO:22); TGTTCACTTT (SEQ ID NO:23);
CAGCAGTCAA (SEQ ID NO:24); AATCACTGCC (SEQ ID NO:25);
TTTCACTGAG (SEQ ID NO:26); ACAGGTGTCA (SEQ ID NO:27);
GCACTAGAAC (SEQ ID NO:28); CAATGCCTAC (SEQ ID NO:29);
AAATTTTAAT (SEQ ID NO:30); TTGATTCAGC (SEQ ID NO:31);
TATGCTGTGA (SEQ ID NO:32); and AAAATGAATG (SEQ ID NO:33).

In some embodiments, a composition contains two or more differently labeled genotyping oligonucleotides for simultaneously probing the identity of nucleotides at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic site.

CTLA4 genotyping oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized genotyping oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized CTLA4 genotyping oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms in multipl genes at the same time.

provide haplotype information on one of the two CTLA4 gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional CTLA4 clones will need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the CTLA4 gene in an individual. In some embodiments, the haplotyping method also comprises identifying the nucleotide at PS2. In a particularly preferred embodiment, the nucleotide at each of PS1-7 is identified.

In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the CTLA4 haplotypes shown in Table 4. This can be accomplished by identifying, for one or both copies of the individual's CTLA4 gene, the phased sequence of nucleotides present at each of PS1-7. The present invention also contemplates that typically only a subset of PS1-7 will need to be directly examined to assign to an individual one or more of the haplotypes shown in Table 4. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysdale, CM et al. 2000 *PNAS* 97:10483-10488; Rieder MJ et al. 1999 *Nature Genetics* 22:59-62). Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stephens, JC 1999, *Mol. Diag.* 4:309-317). Techniques for determining whether any two polymorphic sites are in linkage disequilibrium are well-known in the art (Weir B.S. 1996 *Genetic Data Analysis II*, Sinauer Associates, Inc. Publishers, Sunderland, MA).

In a preferred embodiment, a CTLA4 haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more polymorphic sites selected from the group consisting of PS1, PS3, PS4, PS5, PS6 and PS7 in each copy of the CTLA4 gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-7 in each copy of the CTLA4 gene. When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the CTLA4 gene, or a fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous at that site, while two different nucleotides will be detected if the

individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., *Proc. Natl. Acad. Sci. USA* 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., *Science* 241:1077-1080, 1988).

Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* 89:392-396, 1992).

A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

The genotype or haplotype for the CTLA4 gene of an individual may also be determined by hybridization of a nucleic acid sample containing one or both copies of the gene, or fragment(s) thereof, to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would

contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., *Proc. Natl. Acad. Sci. USA* 82:7575, 1985; Meyers et al., *Science* 230:1242, 1985) and proteins which recognize nucleotide mismatches, such as the E. coli mutS protein (Modrich, P. *Ann. Rev. Genet.* 25:229-253, 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita et al., *Genomics* 5:874-879, 1989; Humphries et al., in Molecular Diagnosis of Genetic Diseases, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., *Nucl. Acids Res.* 18:2699-2706, 1990; Sheffield et al., *Proc. Natl. Acad. Sci. USA* 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455, WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798. Another primer extension method is allele-specific PCR (Ruaño et al., *Nucl. Acids Res.* 17:8392, 1989; Ruaño et al., *Nucl. Acids Res.* 19, 6877-6882, 1991; WO 93/22456; Turki et al., *J. Clin. Invest.* 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in Wallace et al. (WO89/10414).

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by genotyping another polymorphic site that is in linkage disequilibrium with the polymorphic site that is of interest. Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. Genotyping of a polymorphic site in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic site.

In another aspect of the invention, an individual's CTLA4 haplotype pair is predicted from its CTLA4 genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a CTLA4 genotype for the individual at two or more CTLA4 polymorphic sites described herein, enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing CTLA4 haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with th data. In one embodiment, the reference haplotype pairs include the CTLA4 haplotype pairs shown in Table 3.

Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African-descent, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For example, if one wants to have a q% chance of not missing a haplotype that exists in the population at a p% frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by $2n=\log(1-q)/\log(1-p)$ where p and q are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (D.L. Hartl et al., Principles of Population Genomics, Sinauer Associates (Sunderland, MA), 3rd Ed., 1997) postulates that the frequency of finding the haplotype pair H_1 / H_2 is equal to $p_{H-W}(H_1 / H_2) = 2p(H_1)p(H_2)$ if $H_1 \neq H_2$ and $p_{H-W}(H_1 / H_2) = p(H_1)p(H_2)$ if $H_1 = H_2$. A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), single molecule dilution, or allele-specific long-range PCR (Michalotos-Beloin et al., *Nucleic Acids Res.* 24:4841-4843, 1996).

In one embodiment of this method for predicting a CTLA4 haplotype pair for an individual, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is

diseases and other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and a CTLA4 genotype, haplotype, or haplotype pair, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the CTLA4 gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between individual response and CTLA4 genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their CTLA4 genotype or haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the CTLA4 gene giv the most significant contribution to the differences in phenotype. One regression model useful in th invention is described in PCT

Application Serial No. PCT/US00/17540, entitled "Methods for Obtaining and Using Haplotype Data".

A second method for finding correlations between CTLA4 haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2nd Edition (McGraw-Hill, New York, 1991, Ch. 18), standard gradient descent methods (Press et al., *supra*, Ch. 10), or other global or local optimization approaches (see discussion in Judson, *supra*) could also be used. Preferably, the correlation is found using a genetic algorithm approach as described in PCT Application Serial No. PCT/US00/17540.

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the CTLA4 gene. As described in PCT Application Serial No. PCT/US00/17540, ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or variables that can be measured (Fisher and vanBelle, *supra*, Ch. 10).

From the analyses described above, a mathematical model may be readily constructed by the skilled artisan that predicts clinical response as a function of CTLA4 genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the CTLA4 gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the CTLA4 gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying CTLA4 genotype or haplotype that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

In another embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the CTLA4 gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant CTLA4 gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS1, PS3, PS4, PS5, PS6 and PS7, and may also comprise an additional polymorphism of guanine at PS2. Similarly, the nucleotide sequence of a variant fragment of the

CTLA4 gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of the novel polymorphic sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence of the CTLA4 gene, which is defined by haplotype 1, (or other reported CTLA4 sequences) or to portions of the reference sequence (or other reported CTLA4 sequences), except for genotyping oligonucleotides as described above.

The location of a polymorphism in a variant gene or fragment is identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of thymine at PS1, guanine at PS3, cytosine at PS4, thymine at PS5, guanine at PS6 and guanine at PS7. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the CTLA4 gene which is defined by any one of haplotypes 2-8 shown in Table 4 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the CTLA4 gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the novel polymorphic sites described herein. Any particular variant claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art.

CTLA4 isogenes may be isolated using any method that allows separation of the two "copies" of the CTLA4 gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and U.S. Patent No. 5,972,614. Another method, which is described in U.S. Patent No. 5,972,614, uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., *Proc. Natl. Acad. Sci.* 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 1989, *supra*; Ruaño et al., 1991, *supra*; Michalatos-Beloin et al., *supra*).

The invention also provides CTLA4 genome anthologies, which are collections of CTLA4 isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same sex population. A CTLA4 genome anthology may comprise individual CTLA4 isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and the like. Alternatively, two or more groups of the CTLA4 isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred CTLA4 genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 4 below.

An isolated polynucleotid containing a polymorphic variant nucleotide sequence of the invention may b operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded CTLA4 protein in a

prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant CTLA4 sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as E. coli, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect, eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 *Science* 282:1145-1147). Particularly preferred host cells are mammalian cells.

As used herein, a polymorphic variant of a CTLA4 gene fragment comprises at least one novel polymorphism identified herein and has a length of at least 10 nucleotides and may range up to the full length of the gene. Preferably, such fragments are between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

In describing the CTLA4 polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the CTLA4 gene may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are complementary to the sense strand of the CTLA4 genomic variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular CTLA4 protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the CTLA4 isogene encoding that isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular CTLA4 isogene. Expression of a CTLA4 isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA for the isogene. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3′ untranslated region) of the isogene may block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions −10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of CTLA4 mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of CTLA4 mRNA transcribed from a particular isogene.

The oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, the oligonucleotides may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2′ O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

Effect(s) of the polymorphisms identified herein on expression of CTLA4 may be investigated by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant animals, containing a polymorphic variant of the CTLA4 gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into CTLA4 protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired CTLA4 isogene may be introduced into the cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the CTLA4 isogene is introduced into a cell in such a way that it recombines with the endogenous CTLA4 gene present in the cell. Such recombination requires the occurrence of a double recombination event,

thereby resulting in the desired CTLA4 gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the CTLA4 isogene may be introduced include, but are not limited to, continuous culture cells, such as COS, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the CTLA4 isogene. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant nonhuman organisms, i.e., transgenic animals, expressing a variant CTLA4 gene are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the CTLA4 isogenes may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human CTLA4 isogene and producing human CTLA4 protein can be used as biological models for studying diseases related to abnormal CTLA4 expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel CTLA4 isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel CTLA4 isogenes; an antisense oligonucleotide directed against one of the novel CTLA4 isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel CTLA4 isogene described herein. Preferably, the composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel CTLA4 isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or

isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

EXAMPLE 1

This example illustrates examination of various regions of the CTLA4 gene for polymorphic sites.

Amplification of Target Regions

The following target regions were amplified using either the PCR primers represented below or 'tailed' PCR primers, each of which includes a universal sequence forming a noncomplementary 'tail' attached to the 5′ end of each unique sequence in the PCR primer pairs. The universal 'tail' sequence for the forward PCR primers comprises the sequence 5′-TGTAAAACGACGGCCAGT-3′ (SEQ ID NO:34) and the universal 'tail' sequence for the reverse PCR primers comprises the sequence 5′-AGGAAACAGCTATGACCAT-3′ (SEQ ID NO:35). The nucleotide positions of the first and last nucleotide of the forward and reverse primers for each region amplified are presented below and correspond to positions in Figure 1.

PCR Primer Pairs

Fragment No.	Forward Primer	Reverse Primer	PCR Product
Fragment 1	37286-37308	complement of 37909-37889	624 nt
Fragment 2	37608-37630	complement of 38295-38273	688 nt
Fragment 3	37690-37711	complement of 38108-38088	419 nt
Fragment 4	40310-40333	complement of 40820-40798	511 nt
Fragment 5	40466-40488	complement of 41010-40988	545 nt
Fragment 6	41123-41147	complement of 41644-41621	522 nt
Fragment 7	42294-42315	complement of 42918-42896	625 nt

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

Reaction volume	= 10 µl
10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 1 µl
100 ng of human genomic DNA	= 1 µl
10 mM dNTP	= 0.4 µl
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 µl
Forward Primer (10 µM)	= 0.4 µl
Reverse Primer (10 µM)	= 0.4 µl
Water	= 6.6µl

Amplification profile:
97°C - 2 min. 1 cycle

97°C - 15 sec. }

70°C - 45 sec. 10 cycles
72°C - 45 sec.

97°C - 15 sec.
64°C - 45 sec. } 35 cycles
72°C - 45 sec.

Sequencing of PCR Products

The PCR products were purified using a Whatman/Polyfiltronics 100 µl 384 well unifilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 µl of distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were sequenced in both directions using either the primer sets represented below with the positions of their first and last nucleotide corresponding to positions in Figure 1, or the appropriate universal 'tail' sequence as a primer. Reaction products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA Analyzer.

Sequencing Primer Pairs

Fragment No.	Forward Primer	Reverse Primer
Fragment 1	37358-37377	complement of 37882-37863
Fragment 2	37637-37656	complement of 38175-38155
Fragment 3	Tailed Seq	
Fragment 4	40349-40368	complement of 40789-40769
Fragment 5:	Tailed Seq.	
Fragment 6:	41162-41182	complement of 41565-41546
Fragment 7:	42340-42361	complement of 42868-42850

Analysis of Sequences for Polymorphic Sites

Sequence information for a minimum of 80 humans was analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., *Nucleic Acids Res.* 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the CTLA4 gene are listed in Table 2 below.

Table 2. Polymorphic Sites Identified in the CTLA4 Gene

Polymorphic Site Number	PolyId[a]	Nucleotide Position	Reference Allele	Variant Allele	CDS Variant	AA Variant
PS1	743786	37535	C	T		
PS2[R]	743788	37902	A	G	49	T17A
PS3	743794	38038	A	G		
PS4	743804	40867	T	C		
PS5	743808	41547	C	T		
PS6	743812	42460	A	G		
PS7	743814	42508	A	G		

[a]PolyId is a unique identifier assigned to each PS by Genaissance Pharmaceuticals, Inc.
[R]Previously identified in literature

EXAMPLE 2

This exampl illustrates analysis of the CTLA4 polymorphisms identified in the Index

Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in the reference population are shown in Table 3 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 3, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 3 were inferred based on linkage disequilibrium and/or Mendelian inheritance.

Table 3. Genotypes and Haplotype Pairs Observed for CTLA4 Gene

Genotype Number	Polymorphic Sites PS1	PS2	PS3	PS4	PS5	PS6	PS7	HAP Pair	
1	C	A	G	T	C	A	A	2	2
2	C	G	G	T	C	A	A	7	7
3	C	A	G/A	T	C	A	A	2	1
4	C	A	G	T	C	A	A/G	2	3
5	C	A	G	T	C/T	A	A	2	5
6	C	A/G	G	T/C	C	A	A	2	6
7	C/T	A	G	T	C	A	A	2	8
8	C	A	G	T	C	A/G	A	2	4
9	C/T	G/A	G	T	C	A	A	7	8
10	C	A/G	G	T	C	A	A	2	7

The haplotype pairs shown in Table 3 were estimated from the unphased genotypes using a computer-implemented extension of Clark's algorithm (Clark, A.G. 1990 *Mol Bio Evol* 7, 111-122) for assigning haplotypes to unrelated individuals in a population sample, as described in U.S. Provisional Application Serial No. 60/198,340 entitled "A Method and System for Determining Haplotypes from a Collection of Polymorphisms" and the corresponding International Application filed April 18, 2001. In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites. This list of haplotypes is augmented with haplotypes obtained from two families (one three-generation Caucasian family and one two-generation African-American family) and then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals.

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 8 human CTLA4 haplotypes shown in Table 4 below.

Table 4. Haplotypes Identified in the CTLA4 Gene

HAP No.	HAP ID	Polymorphic Sites PS 1	PS 2	PS 3	PS 4	PS 5	PS 6	PS 7
1	745030	C	A	A	T	C	A	A
2	745025	C	A	G	T	C	A	A
3	745029	C	A	G	T	C	A	G
4	745031	C	A	G	T	C	G	A
5	745028	C	A	G	T	T	A	A
6	745032	C	G	G	C	C	A	A
7	745026	C	G	G	T	C	A	A
8	745027	T	A	G	T	C	A	A

Table 5 below shows the percent of chromosomes characterized by a given CTLA4 haplotype for all unrelated individuals in the Index Repository for which haplotype data was obtained. The percent of these unrelated individuals who have a given CTLA4 haplotype pair is shown in Table 6. In Tables 5 and 6, the "Total" column shows this frequency data for all of these unrelated individuals, while the other columns show the frequency data for these unrelated individuals categorized according to their self-identified ethnogeographic origin. Abbreviations used in Tables 5 and 6 are AF = African Descent, AS = Asian, CA = Caucasian, HL = Hispanic-Latino, and NA = Native American.

Table 5. Frequency of Observed CTLA4 Haplotypes In Unrelated Individuals

HAP No.	HAP ID	Total	CA	AF	AS	HL	NA
1	745030	0.61	0.0	2.5	0.0	0.0	0.0
2	745025	48.78	54.76	50.0	25.0	63.89	66.67
3	745029	0.61	2.38	0.0	0.0	0.0	0.0
4	745031	0.61	0.0	2.5	0.0	0.0	0.0
5	745028	0.61	0.0	2.5	0.0	0.0	0.0
6	745032	0.61	0.0	2.5	0.0	0.0	0.0
7	745026	42.07	38.1	40.0	60.0	30.56	33.33
8	745027	6.1	4.76	0.0	15.0	5.56	0.0

Table 6. Frequency of Observed CTLA4 Haplotype Pairs In Unrelated Individuals

HAP1	HAP2	Total	CA	AF	AS	HL	NA
2	2	23.17	23.81	25.0	5.0	38.89	33.33
7	7	19.51	14.29	25.0	35.0	5.56	0.0
2	1	1.22	0.0	5.0	0.0	0.0	0.0
2	3	1.22	4.76	0.0	0.0	0.0	0.0
2	5	1.22	0.0	5.0	0.0	0.0	0.0
2	6	1.22	0.0	5.0	0.0	0.0	0.0
2	8	6.1	9.52	0.0	10.0	5.56	0.0
2	4	1.22	0.0	5.0	0.0	0.0	0.0
7	8	6.1	0.0	0.0	20.0	5.56	0.0
2	7	39.02	47.62	30.0	30.0	44.44	66.67

The size and composition of the Index Repository were chosen to represent the genetic diversity across and within four major population groups comprising the general United States population. For example, as described in Table 1 above, this repository contains approximately equal sample sizes of African-descent, Asian-American, European-American, and Hispanic-Latino population groups. Almost all individuals representing each group had all four grandparents with the same ethnogeographic background. The number of unrelated individuals in the Index Repository provides a sample size that is sufficient to detect SNPs and haplotypes that occur in the general population with high statistical certainty. For instance, a haplotype that occurs with a frequency of 5% in the general population has a probability higher than 99.9% of being observed in a sample of 80 individuals from the general population. Similarly, a haplotype that occurs with a frequency of 10% in a specific population group has a 99% probability of being observed in a sample of 20 individuals from that population group. In addition, the size and composition of the Index Repository means that the relative frequencies determined therein for the haplotypes and haplotype pairs of the CTLA4 gene are likely to be similar to the relative frequencies of these CTLA4 haplotypes and haplotype pairs in the general U.S. population and in the four population groups represented in the Index Repository. The genetic diversity observed for the three Native Americans is presented because it is of scientific interest, but due to the small sample size it lacks statistical significance.

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. A method for haplotyping the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene of an individual, which comprises determining which of the CTLA4 haplotypes shown in the table immediately below defines one copy of the individual's CTLA4 gene, wherein each of the CTLA4 haplotypes comprises a set of polymorphisms whose locations and identities are set forth in the table immediately below:

Haplotype Number[a]									Nucleotide	SEQ ID	Region
1	2	3	4	5	6	7	8	PS No.[b]	Position[c]	NO	Examined[d]
C	C	C	C	C	C	C	T	1	37535	36	37286 - 37909
A	A	A	A	A	G	G	A	2	37902	36	37286 - 37909
A	G	G	G	G	G	G	G	3	38038	36	37608 - 38295
T	T	T	T	T	C	T	T	4	40867	36	40466 - 41010
C	C	C	C	T	C	C	C	5	41547	36	41123 - 41644
A	A	A	G	A	A	A	A	6	42460	36	42294 - 42918
A	A	G	A	A	A	A	A	7	42508	36	42294 - 42918

[a]Alleles for haplotypes are presented 5' to 3' in each column
[b]PS = polymorphic site;
[c]Location of PS within the indicated SEQ ID NO, wherein Nt = nucleotide;
[d]Region examined represents the nucleotide positions defining the start and stop positions of the sequenced region.

2. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-7 on the one copy of the individual's CTLA4 gene.

3. A method for haplotyping the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene of an individual, which comprises determining which of the CTLA4 haplotype pairs shown in the table immediately below defines both copies of the individual's CTLA4 gene, wherein each of the CTLA4 haplotype pairs consists of first and second haplotypes which comprise first and second sets of polymorphisms whose locations and identities are set forth in the table immediately below:

Haplotype Pairs[a]											Nucleotide	SEQ ID	Region
2/2	7/7	2/1	2/3	2/5	2/6	2/8	2/4	7/8	2/7	PS No.[b]	Position[c]	NO	Examined[d]
C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/T	C/C	1	37535	36	37286 - 37909
A/A	G/G	A/A	A/A	A/A	A/G	A/A	A/A	G/A	A/G	2	37902	36	37286 - 37909
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	38038	36	37608 - 38295
T/T	T/T	T/T	T/T	T/T	T/C	T/T	T/T	T/T	T/T	4	40867	36	40466 - 41010
C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	5	41547	36	41123 - 41644
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	6	42460	36	42294 - 42918
A/A	A/A	A/A	A/G	A/A	A/A	A/A	A/A	A/A	A/A	7	42508	36	42294 - 42918

[a]Haplotype pairs are represented as 1st Haplotype/2nd Haplotype; with alleles of each haplotype shown 5' to 3' as 1st Nt/2nd Nt in each column, where Nt = nucleotide;
[b]PS = polymorphic site;
[c]Location of PS within the indicated SEQ ID NO, wherein Nt = nucleotide;
[d]Region examined represents the nucleotide positions defining the start and stop positions of

the sequenced region.

4. The method of claim 3, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-7 on both copies of the individual's CTLA4 gene.

5. A method for genotyping the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene of an individual, comprising determining for the two copies of the CTLA4 gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites (PS) selected from the group consisting of PS1, PS3, PS4, PS5, PS6 and PS7, wherein the one or more PS have the location and alternative alleles shown in SEQ ID NO:36.

6. The method of claim 5, wherein the determining step comprises:

(a) isolating from the individual a nucleic acid mixture comprising both copies of the CTLA4 gene, or a fragment thereof, that are present in the individual;

(b) amplifying from the nucleic acid mixture a target region containing the selected polymorphic site;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

7. The method of claim 5, which comprises determining for the two copies of the CTLA4 gene present in the individual the identity of the nucleotide pair at each of PS1-7.

8. A method for haplotyping the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene of an individual which comprises determining, for one copy of the CTLA4 gene present in the individual, the identity of the nucleotide at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS3, PS4, PS5, PS6 and PS7, wherein the two or more PS have the location and alternative alleles shown in SEQ ID NO:36.

9. The method of claim 8, further comprising determining the identity of the nucleotide at PS2, which has the location and alternative alleles shown in SEQ ID NO:36.

10. The method of claim 8, wherein the determining step comprises:

(a) isolating from the individual a nucleic acid sample containing only one of the two copies of the CTLA4 gene, or a fragment thereof, that is present in the individual;

(b) amplifying from the nucleic acid molecule a target region containing the selected polymorphic site;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized

genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

11. A method for predicting a haplotype pair for the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene of an individual comprising:

(a) identifying a CTLA4 genotype for the individual, wherein the genotype comprises the nucleotide pair at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS3, PS4, PS5, PS6 and PS7, having the location and alternative alleles shown in SEQ ID NO:36;

(b) enumerating all possible haplotype pairs which are consistent with the genotype;

(c) comparing the possible haplotype pairs to the haplotype pair data set forth in the table immediately below; and

(d) assigning a haplotype pair to the individual that is consistent with the data

Haplotype Pairs[a]										PS No.[b]	Nucleotide Position[c]	SEQ ID NO	Region Examined[d]
2/2	7/7	2/1	2/3	2/5	2/6	2/8	2/4	7/8	2/7				
C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/T	C/C	1	37535	36	37286 - 37909
A/A	G/G	A/A	A/A	A/A	A/G	A/A	A/A	G/A	A/G	2	37902	36	37286 - 37909
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	38038	36	37608 - 38295
T/T	T/T	T/T	T/T	T/T	T/C	T/T	T/T	T/T	T/T	4	40867	36	40466 - 41010
C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	5	41547	36	41123 - 41644
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	6	42460	36	42294 - 42918
A/A	A/A	A/A	A/G	A/A	A/A	A/A	A/A	A/A	A/A	7	42508	36	42294 - 42918

[a]Haplotype pairs are represented as 1st Haplotype/2nd Haplotype; with alleles of each haplotype shown 5' to 3' as 1st Nt/2nd Nt in each column, where Nt = nucleotide;
[b]PS = polymorphic site;
[c]Location of PS within the indicated SEQ ID NO, wherein Nt = nucleotide;
[d]Region examined represents the nucleotide positions defining the start and stop positions of the sequenced region.

12. The method of claim 11, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-7, which have the location and alternative alleles shown in SEQ ID NO:36.

13. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene which comprises comparing the frequency of the haplotype or haplotype pair in a population exhibiting the trait with the frequency of the haplotype or haplotype pair in a reference population, wherein the haplotype is selected from haplotypes 1-8 shown in the table presented immediately below, wherein each of the haplotypes comprises a a set of polymorphisms whose locations and identities are set forth in the table immediately below:

Haplotype Number[a]								PS No.[b]	Nucleotide Position[c]	SEQ ID NO	Region Examined[d]
1	2	3	4	5	6	7	8				
C	C	C	C	C	C	C	T	1	37535	36	37286 - 37909
A	A	A	A	A	G	G	A	2	37902	36	37286 - 37909
A	G	G	G	G	G	G	G	3	38038	36	37608 - 38295
T	T	T	T	T	C	T	T	4	40867	36	40466 - 41010
C	C	C	C	T	C	C	C	5	41547	36	41123 - 41644
A	A	A	G	A	A	A	A	6	42460	36	42294 - 42918
A	A	G	A	A	A	A	A	7	42508	36	42294 - 42918

[a]Alleles for haplotypes are presented 5' to 3' in each column
[b]PS = polymorphic site;
[c]Location of PS within the indicated SEQ ID NO, wherein Nt = nucleotide;
[d]Region examined represents the nucleotide positions defining the start and stop positions of the sequenced region,

and the haplotype pair is selected from the haplotype pairs shown in the table immediately below, wherein each of the CTLA4 haplotype pairs consists of first and second haplotypes which comprise first and second sets of polymorphisms whose locations and identities are set forth in the table immediately below:

Haplotype Pairs[a]										PS No.[b]	Nucleotide Position[c]	SEQ ID NO	Region Examined[d]
2/2	7/7	2/1	2/3	2/5	2/6	2/8	2/4	7/8	2/7				
C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/T	C/C	1	37535	36	37286 - 37909
A/A	G/G	A/A	A/A	A/A	A/G	A/A	A/A	G/A	A/G	2	37902	36	37286 - 37909
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	38038	36	37608 - 38295
T/T	T/T	T/T	T/T	T/T	T/C	T/T	T/T	T/T	T/T	4	40867	36	40466 - 41010
C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	5	41547	36	41123 - 41644
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	6	42460	36	42294 - 42918
A/A	A/A	A/A	A/G	A/A	A/A	A/A	A/A	A/A	A/A	7	42508	36	42294 - 42918

[a]Haplotype pairs are represented as 1st Haplotype/2nd Haplotype; with alleles of each haplotype shown 5' to 3' as 1st Nt/2nd Nt in each column, where Nt = nucleotide;
[b]PS = polymorphic site;
[c]Location of PS within the indicated SEQ ID NO, wherein Nt = nucleotide;
[d]Region examined represents the nucleotide positions defining the start and stop positions of the sequenced region,

wherein a higher frequency of the haplotype or haplotype pair in the trait population than in the reference population indicates the trait is associated with the haplotype or haplotype pair.

14. The method of claim 13, wherein the trait is a clinical response to a drug targeting CTLA4.

15. A composition comprising at least one genotyping oligonucleotide for detecting a polymorphism in the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene at a polymorphic site (PS) selected from the group consisting of PS1, PS3, PS4, PS5, PS6 and PS7, having the location and alternative alleles shown in SEQ ID NO:36.

16. The composition of claim 15, wherein the genotyping oligonucleotide is an allele-specific oligonucleotide that specifically hybridizes to an allele f the CTLA4 gene at a region containing

the polymorphic site.

17. The composition of claim 16, wherein the allele-specific oligonucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-9, the complements of SEQ ID NOS:4-9, and SEQ ID NOS:10-21.

18. The composition of claim 15, wherein the genotyping oligonucleotide is a primer-extension oligonucleotide.

19. The composition of claim 18, wherein the primer extension oligonucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:22-33.

20. A kit for genotyping the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene of an individual, which comprises a set of oligonucleotides designed to genotype each of polymorphic sites (PS) PS1, PS3, PS4, PS5, PS6 and PS7, having the location and alternative alleles shown in SEQ ID NO:36.

21. The kit of claim 20, which further comprises oligonucleotides designed to genotype PS2, having the location and alternative alleles shown in SEQ ID NO:36.

22. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

(a) a first nucleotide sequence which comprises a cytotoxic T-lymphocyte-associated protein 4 (CTLA4) isogene, wherein the CTLA4 isogene is selected from the group consisting of isogenes 2-8 shown in the table immediately below and wherein each of the isogenes comprises the regions of the SEQ ID NOS shown in the table immediately below and wherein each of the isogenes 2-8 is further defined by the corresponding set of polymorphisms whose locations and polymorphisms are set forth in the table immediately below

Isogene Number[a]								PS No.[b]	Nucleotide Position[c]	SEQ ID NO	Region Examined[d]
1	2	3	4	5	6	7	8				
C	C	C	C	C	C	C	T	1	37535	36	37286 - 37909
A	A	A	A	A	G	G	A	2	37902	36	37286 - 37909
A	G	G	G	G	G	G	G	3	38038	36	37608 - 38295
T	T	T	T	T	C	T	T	4	40867	36	40466 - 41010
C	C	C	C	T	C	C	C	5	41547	36	41123 - 41644
A	A	A	G	A	A	A	A	6	42460	36	42294 - 42918
A	A	G	A	A	A	A	A	7	42508	36	42294 - 42918

[a]Alleles for isogenes are presented 5' to 3' in each column
[b]PS = polymorphic site;
[c]Location of PS within the indicated SEQ ID NO, wherein Nt = nucleotide;
[d]Region examined represents the nucleotide positions defining the start and stop positions of the sequenced region;

(b) a second nucleotide sequence which comprises a fragment of the first nucleotide sequence, wherein the fragment comprises one or more polymorphisms selected from the group consisting of thymine at PS1, guanine at PS3, cytosine at PS4, thymine at PS5, guanine at

PS6 and guanine at PS7, wherein the selected polymorphism has the location set forth in the table immediately above; and

(c) a third nucleotide sequence which is complementary to the first or second nucleotide sequence.

23. The isolated polynucleotide of claim 22, which is a DNA molecule and comprises both the first and third nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.

24. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 22, wherein the organism expresses a CTLA4 protein encoded by the first nucleotide sequence.

25. The recombinant organism of claim 24, which is a nonhuman transgenic animal.

26. The isolated polynucleotide of claim 22 which consists of the second nucleotide sequence.

27. A computer system for storing and analyzing polymorphism data for the cytotoxic T-lymphocyte-associated protein 4 gene, comprising:

(a) a central processing unit (CPU);

(b) a communication interface;

(c) a display device;

(d) an input device; and

(e) a database containing the polymorphism data;

wherein the polymorphism data comprises the haplotypes set forth in the table immediately below:

Haplotype Number[a]								PS No.[b]	Nucleotide Position[c]	SEQ ID NO	Region Examined[d]
1	2	3	4	5	6	7	8				
C	C	C	C	C	C	C	T	1	37535	36	37286 - 37909
A	A	A	A	A	G	G	A	2	37902	36	37286 - 37909
A	G	G	G	G	G	G	G	3	38038	36	37608 - 38295
T	T	T	T	T	C	T	T	4	40867	36	40466 - 41010
C	C	C	C	T	C	C	C	5	41547	36	41123 - 41644
A	A	A	G	A	A	A	A	6	42460	36	42294 - 42918
A	A	G	A	A	A	A	A	7	42508	36	42294 - 42918

[a]Alleles for haplotypes are presented 5' to 3' in each column
[b]PS = polymorphic site;
[c]Location of PS within the indicated SEQ ID NO, wherein Nt = nucleotide;
[d]Region examined represents the nucleotide positions defining the start and stop positions of the sequenced region;

and the haplotype pairs set forth in the table immediately below:

Haplotype Pairs[a]										PS No.[b]	Nucleotide Position[c]	SEQ ID NO	Region Examined[d]
2/2	7/7	2/1	2/3	2/5	2/6	2/8	2/4	7/8	2/7				
C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/T	C/C	1	37535	36	37286 - 37909
A/A	G/G	A/A	A/A	A/A	A/G	A/A	A/A	G/A	A/G	2	37902	36	37286 - 37909
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	3	38038	36	37608 - 38295
T/T	T/T	T/T	T/T	T/T	T/C	T/T	T/T	T/T	T/T	4	40867	36	40466 - 41010
C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	5	41547	36	41123 - 41644
A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	6	42460	36	42294 - 42918
A/A	A/A	A/A	A/G	A/A	A/A	A/A	A/A	A/A	A/A	7	42508	36	42294 - 42918

[a]Haplotype pairs are represented as 1st Haplotype/2nd Haplotype; with alleles of each haplotype shown 5' to 3' as 1st Nt/2nd Nt in each column, where Nt = nucleotide;
[b]PS = polymorphic site;
[c]Location of PS within the indicated SEQ ID NO, wherein Nt = nucleotide;
[d]Region examined represents the nucleotide positions defining the start and stop positions of the sequenced region.

28. A genome anthology for the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene which comprises CTLA4 isogenes defined by any one of haplotypes 1-8 set forth in the table shown below:

Haplotype Number[a]								PS No.[b]	Nucleotide Position[c]	SEQ ID NO	Region Examined[d]
1	2	3	4	5	6	7	8				
C	C	C	C	C	C	C	T	1	37535	36	37286 - 37909
A	A	A	A	A	G	G	A	2	37902	36	37286 - 37909
A	G	G	G	G	G	G	G	3	38038	36	37608 - 38295
T	T	T	T	T	C	T	T	4	40867	36	40466 - 41010
C	C	C	C	T	C	C	C	5	41547	36	41123 - 41644
A	A	A	G	A	A	A	A	6	42460	36	42294 - 42918
A	A	G	A	A	A	A	A	7	42508	36	42294 - 42918

[a]Alleles for haplotypes are presented 5' to 3' in each column
[b]PS = polymorphic site;
[c]Location of PS within the indicated SEQ ID NO, wherein Nt = nucleotide;
[d]Region examined represents the nucleotide positions defining the start and stop positions of the sequenced region.

POLYMORPHISMS IN THE CTLA4 GENE

ACTCAAATTT CCCCCAGTTT CTCAGTAATC TGCTTTGTAT TATTTTTTTC
AATCAATAAT AACACATTAA ATTCTGTCAT CATGTCTCTA TTTTCTCCTT 33900
TAATCTAGAA TGGTTTTCCA GGCTTTTCCC TTTTTAATCT TTTGTGATGT
TAACATTTTT GAAGAATCTA GGCCAGTTAT TTTGCAGAAT ATGGATTTGT 34000
CTGATTATTT CTTCATGCTT AGATTCAGTT AAAACAATTT TGGCAATAAC
TACACAAAGT TGTTGTGTCC TTCTCAGTGT ATTGTATCTG GAGGTACCCT 34100
ATGTCAGAAC ACTCCATCAC AGTAATCTTC AGTTGGGATC ACTTGGTTAA
AGCAGTATCT GCTAGATTTA TCAATTATCT TTTCCCTATT GTAATTATCA 34200
AGTAATCTGT GGGGTGGTAT ATGTGAATAT ATCTTTTCTC ACCAAGATTC
TTTTACCCAG CATTTTTTTT ACAACCATTT GAATTTTCTT GCTATGACAG 34300
AATTAGTTTC ATAAGAGCAT AATTAAGGAA TCTGGTAAGA CGTCTTTGAA
AGTTCTAATT ACAACCCTCA TCTCTTCTTA TGTTAAACTG TCTTCATTCA 34400
TTTACCTACC TATTTAACAA ATGTTTAAAG AATGCTCGTT TTGTGGCTTG
GCCTGTGACA TTTTCATCAG TGGCTTCATA GTTGCCTAAA CTAATGTCTC 34500
TTCCTGTGCT CCAGAATCCT ATTGGCAACT ATTTTCCATG GATGCCATAA
ACTCAACATA CTCAGATGAA AGCTGTCATT TCACTTTCTC ACCTGCTCTT 34600
CCTGTTATTT ATTCCAGTAA ATTGCACCTC TGTCTTCTCA TTGACCCTAT
AATAAGAAAG GCCTATTTAT TGAGTGGCTA TCTGTTACAA TAGGCCTTTT 34700
ATACATGTTA CATGCATTAT CTCATTTAAT CTTTTCAATA ATCTTGAGAA
TTAGATGCCA TTATCTGGAG GACATTGGGC TTTGAGCCCC AGGTCAACAG 34800
TAAGCAAGTC CCCATGCTAG GATACAAATC CAGTTATCTT GGGCTTCAAA
ATCTTTGCTT TAGCATAGTG ATAAAGGTAT CACTGGCTTC AAATCCTGGC 34900
TCTGCCACTT CCTACGGGTG ATTGGGTAAG TCACCTTAAC TTTCAATGCC
TTGATTTCCT TCTTTATAAA ATGGGAAAAA TGGTAACTCT TGTCTTGTAG 35000
GGTTGTTATG GACTTGAAAG AAACAAAACT TGCTGGGCGC GGTGGCTCAC
GCCTGTAATC CCAGCGCTTT GGGAGGCCGA GGCGGGTAGA TCACCTGAGG 35100
TCAGGAATTT GAGACTAGCC TGGTCAACAT GGCGAAACCC CATCTCTACT
AAAATAATAC AAAAAATTAG CTGGGCGTGG TGGTGGGTGC CTGCAATCCC 35200
AGCTACTTGG GAGGCTGAGG CAGGAGAATC ACTTGAACCC AGGAGGCAGA
GGTTGCAGTG AGCCGAGATC GCGCCACTGC ACTCCAACCT GGGAAACAAG 35300
AGTGAAACTC TGTCTCAAAA AAATAAAAAA GAAAAGAAAG AAAACTTGAC
ATTTGAATAA TTTCTGCAGC ACCACAGTAC TTCTCAAACC ATAATTCTGG 35400
TGTCATTCAT TCTAAACTTC TCCTTCCACT CACTGATTGT GGACTAGTCC
CATAGGGACC TCGTCTTCAC AGCTGTCATT ACCTTAGTTT AGACCAGGGG 35500
TTGGCAAACC ATGGCCCACA GGCCAAATCC AGACTTCACC TATTTTCGTA
AATGAAGTTT TATTGGAATG CACACACACC CATTTGTTTA CATATAATAT 35600
ATGGCTTCTT TCATTCTAAA ATGACAGAGT TGAGTAGTGG CAACAGAGAC
CCCACCGTTT GCAAATCATA ACATATTTAC TATTTTGCCC CCTTCAGAAA 35700
GCTTTCCAAA TCCTGGTGCC TGGCGTATTC CAATAGTCTT CTTCCCAGTC
TTCCTGGTTA CATTTCTCCC TGAACCCATC CTCCCCATCC CTAAAATTTC 35800
CCCCAAATGT GAACTCAGTC ATACCAGTTT GCTCCTTATG TTTCATTGGC
CCTTGCTGCT AAGAGCATCC GCTTGCACCT TCTGCTCATC CCCAGACAAG 35900
CTTTGTCCTG TGACCATAAT GAACTCTTCA TGCCGTTTCC AACTTTAGCC
CATGTTATTC TTCTTGTCTG AATATCCACC CTTTTCTCTG TTCTCAATAA 36000
TAAGTTCAGG CTTTTCGTCT TCTGAGAAGC CCTTTCTGAC TTCCACAGGC
TGAACCACTG GCTTCTGCTC CTCTACATAA TACTTCAATT CCAGCATTGA 36100
TCTCACTCTA TCATGATCAT GGGTTTAGCT GTCTGTCCCT GCCACTGCTG
TGTGTTCCTC TTGAGGGCAG GAACATTTGT TTTTCACTTT TTAAAAAACC 36200
TCTGTTGCCC AGTCTGGCAT TAGGAAGTGC CCATTAGGTT GTTATTGCTT
GTTGGCGCTT GAGCTGGGGC TTGAAGGTTT CTATAATGTG TAGCAGTGTA 36300
TAGAAAACAG GCAGGTCAGA AAAGGCTTCT GTGCATCACA CCAACATGGC

FIGURE 1A

```
ACATGTATAC ATATGTAACA AATCTGCATG TTGTGCACAT GTACCCTAAA      36400
ACTTAAAGTA TAATAATAAT AAAATTTTAA AAAAAAAAAG AAGAGGCTTC
CTGGAGGAGA TGACAGCTGA GCTAAGTCCT GGAGGATGAG AAGGAGTATA      36500
AAATAAGATA ATAGGAGAAA AAAGGCAGTA GGAACAGCAT GGGTAAAGGT
GATGAGGCCT GAAAGAGGCA CGTGGAAGGA AAGACAAATG CAGGAAGGGG      36600
GAATGGGAGG GAATGCTGGG GTACAGGCCA AAGAGGGAGG CATTCGGGGA
GTGTTCTTTT AAACATCCTT TGGTGTGCTA ATGTGTGGTC ATTGGGAAAC      36700
CATGGACCGT TTTTTTTTGG CAAACCTCAT ATCCCCTGTT ACAACTGTCT
GTTTGCATGT CAGCCTTGTA AAAGCCCCTT AAGGTATCAA CTATGTTTTT      36800
GTTTTGTCAT CATTCAATCC TAAGTGCACA GAATTCCGGG CATATTACAG
GTTCCCCATG AATGTTTCTT TCTTTATTAA AATGTATGAA AACTCTCCAG      36900
ATTTAAGGAA GGTCCTCAAT GTTTCAAATT CTTTTTGTTA GATCATTGGT
CCTGTCTACA GCTGTCACAA ATTTAAGGAC TCTGGTTATA TTTAATCTTC      37000
ACTTTTGAAT TTTCTGCTTG AAAAATTTGT ATTAGAAAAA AAAGTCTATC
CTTTTATGGA CGGCTCTAAT CTCTTGAATC ATTTGGGTTG GCTTTTCTTT      37100
GGACCTTCTT CAACTCTGTT TTGTCTCTGT TGAGTTAAGG CTTTTAAGAA
CACCTGAATT CTTTCCTTCT GCAAAACCAG AGGCAGCTTC TTTTCCGCCT      37200
ATTTTCAGTT TATTTCTTGT GATTTTAGTT TTTTTCTCTT AACCAAATGC
TAAATGGATT TAGGAGAAAT AAACTTATTT GTAAAGCTGT CAAGGGACCA      37300
TTAGAAGGAT GGTGCTTCAC AGATAGAATA CAGTTTTTAT TAATGATGCC
TAGACAAATC CTGCCATTAG CCCAAGGGCT CAGAAAGTTA GCAGCCTAGT      37400
AGTTTTGGAG TTGTCAATGA AATGAATTGG ACTGGATGGT TAAGGATGCC
CAGAAGATTG AATAAAATTG GGATTTAGGA GGACCCTTGT ACTCCAGGAA      37500
ATTCTCCAAG TCTCCACTTA GTTATCCAGA TCCTCAAAGT GAACATGAAG
                                        T
CTTCAGTTTC AAATTGAATA CATTTTCCAT CCATGGATTG GCTTGTTTTG      37600
TTCAGTTGAG TGCTTGAGGT TGTCTTTTCG ACGTAACAGC TAAACCCACG
GCTTCCTTTC TCGTAAAACC AAAACAAAAA GGCTTTCTAT TCAAGTGCCT      37700
TCTGTGTGTG CACATGTGTA ATACATATCT GGGATCAAAG CTATCTATAT
AAAGTCCTTG ATTCTGTGTG GGTTCAAACA CATTTCAAAG CTTCAGGATC      37800
CTGAAAGGTT TTGCTCTACT TCCTGAAGAC CTGAACACCG CTCCCATAAA
GCCATGGCTT GCCTTGGATT TCAGCGGCAC AAGGCTCAGC TGAACCTGGC      37900
     [exon 1: 37854..
TACCAGGACC TGGCCCTGCA CTCTCCTGTT TTTTCTTCTC TTCATCCCTG
 G
TCTTCTGCAA AGGTGAGTGA GACTTTTGGA GCATGAAGAT GGAGGAGGTG      38000
                ..37961]
TTTCTCCTAC CTGGGTTTCA TTTGTTTCAG CAGTCAAAGG CAGTGATTTA
                                          G
TAGCAAAGCC AGAAGTTAAA GGTAAAACTC CAATCTGGCT TGGCTGGCTC      38100
TGTATTCCAG GGCCAGCAGG GAGCAGTTGG GCGGCACGAA ATAAGGCAAA
GAGATAGCTC AGAACAGAGC GCCAGGTATT TAGTAGGGGC TTCATGAATG      38200
CATGTGAGTT GGTTTAGTAG AGAGACACAG GCAATTTCAG ACCCTTCTAT
GAGACTGGAA GTGATTTAAG AGGGAAAGGA TAGCCATAGT CCTGAATACA      38300
TTTGAGCTGG GTTTCAGGAT GAGCTCACAA GTTCCTTTAA AAAAAATTGA
CTTAAGCAAA TCCTGGGAAG AGTTTTTTTG CTATACAATT CAAGGTTTTA      38400
AGGTCCTCGG ATTCATATAC TTTATAAATG AATTAGCCAG CTTGTTTAAA
ATGTAGGGAA ATTGTGGGAA GAATGCCTTC TTTACTTAAT TCAAGGTTTT      38500
AAGGTTCTCT TAATCAATTC TACTAGCTAA TTAGCCAATT ATTTAAAAAT
AAAAGTTTGA AATTGCCAAA AAAAAAAGAC AAGGAAAAGG AAAGAAAGAA      38600
AGCCACCAGT CTGTTTGGCA TACAATACTT AATTGTTGCC TGACCTACGT
GTGGGTTTCA GATGCAGATC CTCAGTTTTC AGCTCTTCAG AGACTGACAC      38700
CAGGTTTGTT ACACGGCTTA AAATGATGAG TATATCCATT GAATCTCAAC
```

FIGURE 1B

```
CTTATCTCTC TCTAGACCTT CTTGGTTAAG AAACCATGTA GTTTGTATGA      38800
AGTAGGTACT CAAAAGATAT TTGATGATTT AATTTTTACT GGAGAAGAAA
TATTCATATA TGTTTTCTTA TTTTTACATG TTTTAAATAT GTAAAGATTA      38900
AATAAACACT CTTAGAAGTA TTTAAATTTC CTAAAGTAAA TTTATCTCAA
CCAGTAACAG GACCCTCCCA ATACTGGAAA GTTGAGTGTG ACCGCATTTA      39000
GTGGTGATGA GTGTGAGCTT GCTTGGGGAG AGGGCAGGAC ATTTAGGATT
TCTTAAGCTT AGAGTCAATA CAATAAAGAT TATTGAGTGC TCACTTGGGT      39100
GGGCTATAAT CACTGCTCAC AGGAGTTCAT GAACCACAAG TAAAAGAGTG
AGGAGATATG ATTAGCTCAC AAATAACTTT AATACAGAGC AGAAAGTAAT      39200
GAACTACTGC AATGGAGTTA TCACAGTGCT AAGGATGCTC AGAGGGCATC
TCTGATAGGC AGAGGTGAGG GTTAGGGAAG GAAGCTGTAG TCTAGCTAGC      39300
TAGAGCTGCT GGAATAGACA TGACAATGGC TGCTGCAAAC TGTTTTCTCT
TCTGAGGACA GATGTCCCGT GCAAGTGGCT TGGTGGAAGG GACTAGTGTC      39400
TCTAATATAG GGTGATTTAT AAGCAGGAAA GTGTGTCCTA GAAATTCAGA
CCAGAGTGAT AGATTGGAAT TGGATCATGG GGGACTCATT GAATGTTATT      39500
TATTGTATTT GTTTTTGCGA TCAGTGTTAG TAAAGTGTCA AAGGGATTGA
GCAGATGAGT GACATCATGC AACACAAGTT TTGAGTTTCA CTTGTCAGAC      39600
TGACTGGAGA GGGGCCTGGT TAGTTACAGG AAGGTAATTT GGCATGCAGC
CACTATTTTT GAGTTGATGC AAGCCTCTCT GTATGGAGAG CTGGTCTCCT      39700
TTATCCTGTG GGAAAAGAGA ACAAAGGAGC ATGGGAGTGT TCAAGGGAAG
GAGAAATAAA GGGCAGAGAG GCAGCGGTGG TGTCAGGGGA AGCCCACAGG      39800
AGTTAACAGC AGGGTTGCCT CAACCTAGAG AGGAAGCGAC CTGGTGCCCT
CGGCTCTGTG GCTTCCTTCA TCTAACAACA TCTTCCACTC TACAACAATG      39900
CCAGGGAAGG CGGAGGCTGG TACAGTGCAT CAAGACACAG CTACTCCTGG
GTGACAGAGG TTCAGGGCCA GCTCACTAAG TAGGCAGAAG TTTTTGACAT      40000
ATACTTTGAG AGATAAAGCA AGATTCTGTA CCTCAACCTT CAGAATTTCC
CCTACCACTC ATTATAGTTC CGGAGCTATA TAGCTCCTAT CATTCTATCA      40100
TAACCTTAGA ATACCAGAGA ACATATCATC TCATCTAATT ATCTCTTACT
ATATGTGAAA AAAATGAAGG ACATGGGGGA AGTGTGACTT GCCCCAAATC      40200
ACATATTTCA TGGTAGAGCC AGGTCTTCTG TTTGTCATAT CAGTGTTCTT
CCTGCCACAA CCATCTTGAA GAATCTATTT CTCAGTAAGA AAATATCTTT      40300
ATGGAGAGTA GCTGGAAAAC AGTTGAGAGA TGGAGGGGAG GCTGGGGGTG
TGGAGAGGGG AAGGGGTAAG TGATAGATTC GTTGAAGGGG GGAGAAAAGG      40400
CCGTGGGGAT GAAGCTAGAA GGCAGAAGGG CTTGCCTGGG CTTGGCCATG
AAGGAGCATG AGTTCACTGA GTTCCCTTTG GCTTTTCCAT GCTAGCAATG      40500
     [exon 2: 40495..
CACGTGGCCC AGCCTGCTGT GGTACTGGCC AGCAGCCGAG GCATCGCCAG
CTTTGTGTGT GAGTATGCAT CTCCAGGCAA AGCCACTGAG GTCCGGGTGA      40600
CAGTGCTTCG GCAGGCTGAC AGCCAGGTGA CTGAAGTCTG TGCGGCAACC
TACATGATGG GGAATGAGTT GACCTTCCTA GATGATTCCA TCTGCACGGG      40700
CACCTCCAGT GGAAATCAAG TGAACCTCAC TATCCAAGGA CTGAGGGCCA
TGGACACGGG ACTCTACATC TGCAAGGTGG AGCTCATGTA CCCACCGCCA      40800
TACTACCTGG GCATAGGCAA CGGAACCCAG ATTTATGTAA TTGGTGAGCA
                ..40843]
AAGCCATTTC ACTGAGTTGA CACCTGTTGC ATTGCAGTCT TCTATGCACA      40900
                C
AAAACAGTTT TGTTCCTTAA TTTCAGGAGG TTTACTTTTA GGACTGTGGA
CATTCTCTTT AAGAGTTCTG TACCACATGG TAGCCTTGCT TATTGTGGGT      41000
GGCAACCTTA ATAGCATTCT GACTGTAAAA TAAAATGATT TGGGGAAGTT
GGGGCTCTCG CTCTGGAGTG CTAACCATCA TGACGTTTGA TCTGTACTTT      41100
TGATATGATA TGATGCTCCT GGGGAAGTAG TCCCAAATAG CCAAACCTAT
TGGTGGGCTA CCCATGCAAT TTAGGGGTGG ACCTCAAGGC CTGGAAGCTC      41200
TAATGTCCTT TTTTCACCAA TGTTGGGGAG TAGAGCCCTA GAGTTTAAAA
```

FIGURE 1C

POLYMORPHISMS IN THE CODING SEQUENCE OF CTLA4

```
ATGGCTTGCC TTGGATTTCA GCGGCACAAG GCTCAGCTGA ACCTGGCTAC
                                                      G
CAGGACCTGG CCCTGCACTC TCCTGTTTTT TCTTCTCTTC ATCCCTGTCT      100
TCTGCAAAGC AATGCACGTG GCCCAGCCTG CTGTGGTACT GGCCAGCAGC
CGAGGCATCG CCAGCTTTGT GTGTGAGTAT GCATCTCCAG GCAAAGCCAC      200
TGAGGTCCGG GTGACAGTGC TTCGGCAGGC TGACAGCCAG GTGACTGAAG
TCTGTGCGGC AACCTACATG ATGGGGAATG AGTTGACCTT CCTAGATGAT      300
TCCATCTGCA CGGGCACCTC CAGTGGAAAT CAAGTGAACC TCACTATCCA
AGGACTGAGG GCCATGGACA CGGGACTCTA CATCTGCAAG GTGGAGCTCA      400
TGTACCCACC GCCATACTAC CTGGGCATAG GCAACGGAAC CCAGATTTAT
GTAATTGATC CAGAACCGTG CCCAGATTCT GACTTCCTCC TCTGGATCCT      500
TGCAGCAGTT AGTTCGGGGT TGTTTTTTTA TAGCTTTCTC CTCACAGCTG
TTTCTTTGAG CAAAATGCTA AAGAAAAGAA GCCCTCTTAC AACAGGGGTC      600
TATGTGAAAA TGCCCCCAAC AGAGCCAGAA TGTGAAAAGC AATTTCAGCC
TTATTTTATT CCCATCAATT GA                                    672
```

FIGURE 2

SEQUENCE LISTING

<110> Genaissance Pharmaceuticals, Inc.
Chew, Anne
Choi, Julie Y.
Messer, Chad

<120> Haplotypes of the CTLA4 Gene

<130> MWH-0537PCT CTLA4

<140> TBA
<141> 2001-05-23

<150> 60/206,353
<151> 2000-05-23

<160> 36

<170> PatentIn Ver. 2.1

<210> 1
<211> 9336
<212> DNA
<213> Homo sapien

<400> 1
actcaaattt cccccagttt ctcagtaatc tgctttgtat tatttttttc aatcaataat 60
aacacattaa attctgtcat catgtctcta ttttctcctt taatctagaa tggttttcca 120
ggcttttccc tttttaatct tttgtgatgt taacattttt gaagaatcta ggccagttat 180
tttgcagaat atggatttgt ctgattattt cttcatgctt agattcagtt aaaacaattt 240
tggcaataac tacacaaagt tgttgtgtcc ttctcagtgt attgtatctg gaggtaccct 300
atgtcagaac actccatcac agtaatcttc agttgggatc acttggttaa agcagtatct 360
gctagattta tcaattatct tttccctatt gtaattatca agtaatctgt ggggtggtat 420
atgtgaatat atcttttctc accaagattc ttttacccag catttttttt acaaccattt 480
gaattttctt gctatgacag aattagtttc ataagagcat aattaaggaa tctggtaaga 540
cgtctttgaa agttctaatt acaaccctca tctcttctta tgttaaactg tcttcattca 600
tttacctacc tatttaacaa atgtttaaag aatgctcgtt ttgtggcttg gcctgtgaca 660
ttttcatcag tggcttcata gttgcctaaa ctaatgtctc ttcctgtgct ccagaatcct 720
attggcaact attttccatg gatgccataa actcaacata ctcagatgaa agctgtcatt 780
tcactttctc acctgctctt cctgttattt attccagtaa attgcacctc tgtcttctca 840
ttgaccctat aataagaaag gcctatttat tgagtggcta tctgttacaa taggcctttt 900
atacatgtta catgcattat ctcatttaat cttttcaata atcttgagaa ttagatgcca 960
ttatctggag gacattgggc tttgagcccc aggtcaacag taagcaagtc cccatgctag 1020
gatacaaatc cagttatctt gggcttcaaa atctttgctt tagcatagtg ataaaggtat 1080
cactggcttc aaatcctggc tctgccactt cctacgggtg attgggtaag tcaccttaac 1140
tttcaatgcc ttgatttcct tctttataaa atgggaaaaa tggtaactct tgtcttgtag 1200
ggttgttatg gacttgaaag aaacaaaact tgctgggcgc ggtggctcac gcctgtaatc 1260

ccagcgcttt gggaggccga ggcgggtaga tcacctgagg tcaggaattt gagactagcc 1320
tggtcaacat ggcgaaaccc catctctact aaaataatac aaaaaattag ctgggcgtgg 1380
tggtgggtgc ctgcaatccc agctacttgg gaggctgagg caggagaatc acttgaaccc 1440
aggaggcaga ggttgcagtg agccgagatc gcgccactgc actccaacct gggaaacaag 1500
agtgaaactc tgtctcaaaa aaataaaaaa gaaaagaaag aaaacttgac atttgaataa 1560
tttctgcagc accacagtac ttctcaaacc ataattctgg tgtcattcat tctaaacttc 1620
tccttccact cactgattgt ggactagtcc catagggacc tcgtcttcac agctgtcatt 1680
accttagttt agaccagggg ttggcaaacc atggcccaca ggccaaatcc agacttcacc 1740
tattttcgta aatgaagttt tattggaatg cacacacacc catttgttta catataatat 1800
atggcttctt tcattctaaa atgacagagt tgagtagtgg caacagagac cccaccgttt 1860
gcaaatcata acatatttac tattttgccc ccttcagaaa gctttccaaa tcctggtgcc 1920
tggcgtattc caatagtctt cttcccagtc ttcctggtta catttctccc tgaacccatc 1980
ctccccatcc ctaaaatttc ccccaaatgt gaactcagtc ataccagttt gctccttatg 2040
tttcattggc ccttgctgct aagagcatcc gcttgcacct tctgctcatc cccagacaag 2100
ctttgtcctg tgaccataat gaactcttca tgccgtttcc aactttagcc catgttattc 2160
ttcttgtctg aatatccacc cttttctctg ttctcaataa taagttcagg cttttcgtct 2220
tctgagaagc cctttctgac ttccacaggc tgaaccactg gcttctgctc ctctacataa 2280
tacttcaatt ccagcattga tctcactcta tcatgatcat gggtttagct gtctgtccct 2340
gccactgctg tgtgttcctc ttgagggcag gaacatttgt ttttcacttt ttaaaaaacc 2400
tctgttgccc agtctggcat taggaagtgc ccattaggtt gttattgctt gttggcgctt 2460
gagctggggc ttgaaggttt ctataatgtg tagcagtgta tagaaaacag gcaggtcaga 2520
aaaggcttct gtgcatcaca ccaacatggc acatgtatac atatgtaaca aatctgcatg 2580
ttgtgcacat gtaccctaaa acttaaagta taataataat aaaatttttaa aaaaaaaaag 2640
aagaggcttc ctggaggaga tgacagctga gctaagtcct ggaggatgag aaggagtata 2700
aaataagata ataggagaaa aaaggcagta ggaacagcat gggtaaaggt gatgaggcct 2760
gaaagaggca cgtggaagga aagacaaatg caggaagggg gaatgggagg gaatgctggg 2820
gtacaggcca aagagggagg cattcgggga gtgttctttt aaacatcctt tggtgtgcta 2880
atgtgtggtc attgggaaac catggaccgt ttttttttgg caaacctcat atcccctgtt 2940
acaactgtct gtttgcatgt cagccttgta aaagcccctt aaggtatcaa ctatgttttt 3000
gttttgtcat cattcaatcc taagtgcaca gaattccggg catattacag gttccccatg 3060
aatgtttctt tctttattaa aatgtatgaa aactctccag atttaaggaa ggtcctcaat 3120
gtttcaaatt ctttttgtta gatcattggt cctgtctaca gctgtcacaa atttaaggac 3180
tctggttata tttaatcttc acttttgaat tttctgcttg aaaaatttgt attagaaaaa 3240
aaagtctatc cttttatgga cggctctaat ctcttgaatc atttgggttg gcttttcttt 3300
ggaccttctt caactctgtt ttgtctctgt tgagttaagg cttttaagaa cacctgaatt 3360
ctttccttct gcaaaaccag aggcagcttc ttttccgcct attttcagtt tatttcttgt 3420
gattttagtt tttttctctt aaccaaatgc taaatggatt taggagaaat aaacttattt 3480
gtaaagctgt caagggacca ttagaaggat ggtgcttcac agatagaata cagtttttat 3540
taatgatgcc tagacaaatc ctgccattag cccaagggct cagaaagtta gcagcctagt 3600
agttttggag ttgtcaatga aatgaattgg actggatggt taaggatgcc cagaagattg 3660
aataaaattg ggatttagga ggacccttgt actccaggaa attctccaag tctccactta 3720
gttatccaga tcctcaaagt gaacatgaag cttcagtttc aaattgaata cattttccat 3780
ccatggattg gcttgttttg ttcagttgag tgcttgaggt tgtcttttcg acgtaacagc 3840
taaacccacg gcttcctttc tcgtaaaacc aaaacaaaaa ggctttctat tcaagtgcct 3900
tctgtgtgtg cacatgtgta atacatatct gggatcaaag ctatctatat aaagtccttg 3960
attctgtgtg ggttcaaaca catttcaaag cttcaggatc ctgaaaggtt ttgctctact 4020
tcctgaagac ctgaacaccg ctcccataaa gccatggctt gccttggatt tcagcggcac 4080
aaggctcagc tgaacctggc taccaggacc tggccctgca ctctcctgtt ttttcttctc 4140

ttcatccctg tcttctgcaa aggtgagtga gacttttgga gcatgaagat ggaggaggtg 4200
tttctcctac ctgggtttca tttgtttcag cagtcaaagg cagtgattta tagcaaagcc 4260
agaagttaaa ggtaaaactc caatctggct tggctggctc tgtattccag ggccagcagg 4320
gagcagttgg gcggcacgaa ataaggcaaa gagatagctc agaacagagc gccaggtatt 4380
tagtaggggc ttcatgaatg catgtgagtt ggtttagtag agagacacag gcaatttcag 4440
acccttctat gagactggaa gtgatttaag agggaaagga tagccatagt cctgaataca 4500
tttgagctgg gtttcaggat gagctcacaa gttcctttaa aaaaaattga cttaagcaaa 4560
tcctgggaag agtttttttg ctatacaatt caaggtttta aggtcctcgg attcatatac 4620
tttataaatg aattagccag cttgtttaaa atgtagggaa attgtgggaa gaatgccttc 4680
tttacttaat tcaaggtttt aaggttctct taatcaattc tactagctaa ttagccaatt 4740
atttaaaaat aaaagtttga aattgccaaa aaaaaaagac aaggaaaagg aaagaaagaa 4800
agccaccagt ctgtttggca tacaatactt aattgttgcc tgacctacgt gtgggtttca 4860
gatgcagatc ctcagttttc agctcttcag agactgacac caggtttgtt acacggctta 4920
aaatgatgag tatatccatt gaatctcaac cttatctctc tctagacctt cttggttaag 4980
aaaccatgta gtttgtatga agtaggtact caaaagatat ttgatgattt aatttttact 5040
ggagaagaaa tattcatata tgttttctta tttttacatg ttttaaatat gtaaagatta 5100
aataaacact cttagaagta tttaaatttc ctaaagtaaa tttatctcaa ccagtaacag 5160
gaccctccca atactggaaa gttgagtgtg accgcattta gtggtgatga gtgtgagctt 5220
gcttggggag agggcaggac atttaggatt tcttaagctt agagtcaata caataaagat 5280
tattgagtgc tcacttgggt gggctataat cactgctcac aggagttcat gaaccacaag 5340
taaaagagtg aggagatatg attagctcac aaataacttt aatacagagc agaaagtaat 5400
gaactactgc aatggagtta tcacagtgct aaggatgctc agagggcatc tctgataggc 5460
agaggtgagg gttagggaag gaagctgtag tctagctagc tagagctgct ggaatagaca 5520
tgacaatggc tgctgcaaac tgttttctct tctgaggaca gatgtcccgt gcaagtggct 5580
tggtggaagg gactagtgtc tctaatatag ggtgatttat aagcaggaaa gtgtgtccta 5640
gaaattcaga ccagagtgat agattggaat tggatcatgg gggactcatt gaatgttatt 5700
tattgtattt gtttttgcga tcagtgttag taaagtgtca aagggattga gcagatgagt 5760
gacatcatgc aacacaagtt ttgagtttca cttgtcagac tgactggaga ggggcctggt 5820
tagttacagg aaggtaattt ggcatgcagc cactattttt gagttgatgc aagcctctct 5880
gtatggagag ctggtctcct ttatcctgtg ggaaaagaga acaaaggagc atgggagtgt 5940
tcaagggaag gagaaataaa gggcagagag gcagcggtgg tgtcagggga agcccacagg 6000
agttaacagc agggttgcct caacctagag aggaagcgac ctggtgccct cggctctgtg 6060
gcttccttca tctaacaaca tcttccactc tacaacaatg ccagggaagg cggaggctgg 6120
tacagtgcat caagacacag ctactcctgg gtgacagagg ttcagggcca gctcactaag 6180
taggcagaag tttttgacat atactttgag agataaagca agattctgta cctcaacctt 6240
cagaatttcc cctaccactc attatagttc cggagctata tagctcctat cattctatca 6300
taaccttaga ataccagaga acatatcatc tcatctaatt atctcttact atatgtgaaa 6360
aaaatgaagg acatgggggа agtgtgactt gccccaaatc acatatttca tggtagagcc 6420
aggtcttctg tttgtcatat cagtgttctt cctgccacaa ccatcttgaa gaatctattt 6480
ctcagtaaga aaatatcttt atggagagta gctggaaaac agttgagaga tggaggggag 6540
gctgggggtg tggagagggg aaggggtaag tgatagattc gttgaagggg ggagaaaagg 6600
ccgtggggat gaagctagaa ggcagaaggg cttgcctggg cttggccatg aaggagcatg 6660
agttcactga gttccctttg gcttttccat gctagcaatg cacgtggccc agcctgctgt 6720
ggtactggcc agcagccgag gcatcgccag ctttgtgtgt gagtatgcat ctccaggcaa 6780
agccactgag gtccgggtga cagtgcttcg gcaggctgac agccaggtga ctgaagtctg 6840
tgcggcaacc tacatgatgg ggaatgagtt gaccttccta gatgattcca tctgcacggg 6900
cacctccagt ggaaatcaag tgaacctcac tatccaagga ctgagggcca tggacacggg 6960
actctacatc tgcaaggtgg agctcatgta cccaccgcca tactacctgg gcataggcaa 7020

cggaacccag atttatgtaa ttggtgagca aagccatttc actgagttga cacctgttgc 7080
attgcagtct tctatgcaca aaaacagttt tgttccttaa tttcaggagg tttacttttа 7140
ggactgtgga cattctcttt aagagttctg taccacatgg tagccttgct tattgtgggt 7200
ggcaaccttа atagcattct gactgtaaaa taaaatgatt tggggaagtt ggggctctcg 7260
ctctggagtg ctaaccatca tgacgtttga tctgtacttt tgatatgata tgatgctcct 7320
ggggaagtag tcccaaatag ccaaacctat tggtgggcta cccatgcaat ttaggggtgg 7380
acctcaaggc ctggaagctc taatgtcctt ttttcaccaa tgttggggag tagagcccta 7440
gagtttaaaa ctgtctcagg gaggctctgc tttgttttct gttgcagatc cagaaccgtg 7500
cccagattct gacttcctcc tctggatcct tgcagcagtt agttcggggt tgtttttttа 7560
tagctttctc ctcacagctg tttctttgag caaaatggtg agtgtggtgc tgatggtgca 7620
ccatgtctga tggggatacc tttagtggta tcaactggcc aaaagatgat gttgagttta 7680
gtgttcttga gatgagatga ggcaataaat gaagaggaag gacagtggta aagaacgcac 7740
tagaaccgta ggcattggca tttgaggttt cagaatgact aatattttag atgaatttgt 7800
ttgacattga atgttcatgt gcttctgagc agggtttcaa tttgagtaac cgttgcaata 7860
acatggggca gctgttttgc tctttgtctt catgacaact gtacttaagc taacagccct 7920
gaaacatgag attaggctgg gcagaatgct gctagagagg accacttgga tggtctttat 7980
tctccttctc catgtccctc tccatcacct ggaagtcacc tctgggtgcc actctggtgc 8040
cttccttgtc gaagctgtag ctgctcacat gacacctatc cctgttatcc agtttgcttg 8100
actgggacgt tttgccttcc ccttcagcca ggaagtgaaa gtcccagttt ttatttatca 8160
caggtgttgg tattggtggt agaagaggta gaattatgga atcaggcctc ctgtcaggat 8220
ttctttttga cagtccctct cagacacctc tgcctaaggc cagctttgcc attacaaact 8280
ctcccttctc cctctctccc ttcttctctt cctcttcctt cttctcgctc tttctctctc 8340
tctctttctc cctctctgtc tcttatacac atacacaaag atatactcta ttccaacatc 8400
ctctacccaa cctgacagag atgtcctttg ctgtaggttc agcagtgggg atgagaaata 8460
cagctctcaa acaggataac taaagcttat tatcttatca agcttgttcc cttgcagaca 8520
agattgatca attatcatag gctttctggg tgttctttct gaagctttct caaagtctct 8580
ttctcctatc ttccattcaa ggcaaatgat tgccatttaa catcaaaatc acagttattt 8640
atctaaaata aatttaata gctgaatcaa gaaaatctcc tgaggtttat aattctgtat 8700
gctgtgaaca ttcattttta accagctagg gacccaatat gtgttgagtt ctattatggt 8760
tagaagtggc ttccgtattc ctcagtagta attactgttt ctttttgtgt ttgacagcta 8820
aagaaaagaa gccctcttac aacaggggtc tatgtgaaaa tgcccccaac agagccagaa 8880
tgtgaaaagc aatttcagcc ttatttttatt cccatcaatt gagaaaccat tatgaagaag 8940
agagtccata tttcaatttc caagagctga ggcaattcta acttttttgc tatccagcta 9000
tttttatttg tttgtgcatt tgggggggaat tcatctctct ttaatataaa gttggatgcg 9060
gaacccaaat tacgtgtact acaatttaaa gcaaaggagt agaaagacag agctgggatg 9120
tttctgtcac atcagctcca ctttcagtga aagcatcact tgggattaat atggggatgc 9180
agcattatga tgtgggtcaa ggaattaagt tagggaatgg cacagcccaa agaaggaaag 9240
gcagggagcg agggagaaga ctatattgta cacaccttat atttacgtat gagacgttta 9300
tagccgaaat gatctgttca agttagatgt tatgcc 9336

<210> 2
<211> 672
<212> DNA
<213> Homo sapien

<400> 2
atggcttgcc ttggatttca gcggcacaag gctcagctga acctggctac caggacctgg 60

ccctgcactc tcctgttttt tcttctcttc atccctgtct tctgcaaagc aatgcacgtg 120
gcccagcctg ctgtggtact ggccagcagc cgaggcatcg ccagctttgt gtgtgagtat 180
gcatctccag gcaaagccac tgaggtccgg gtgacagtgc ttcggcaggc tgacagccag 240
gtgactgaag tctgtgcggc aacctacatg atggggaatg agttgacctt cctagatgat 300
tccatctgca cgggcacctc cagtggaaat caagtgaacc tcactatcca aggactgagg 360
gccatggaca cgggactcta catctgcaag gtggagctca tgtacccacc gccatactac 420
ctgggcatag gcaacggaac ccagatttat gtaattgatc cagaaccgtg cccagattct 480
gacttcctcc tctggatcct tgcagcagtt agttcggggt tgtttttta tagctttctc 540
ctcacagctg tttctttgag caaaatgcta aagaaaagaa gccctcttac aacaggggtc 600
tatgtgaaaa tgccccaac agagccagaa tgtgaaaagc aatttcagcc ttatttatt 660
cccatcaatt ga 672

<210> 3
<211> 223
<212> PRT
<213> Homo sapien

<400> 3
Met Ala Cys Leu Gly Phe Gln Arg His Lys Ala Gln Leu Asn Leu Ala
1 5 10 15

Thr Arg Thr Trp Pro Cys Thr Leu Leu Phe Phe Leu Leu Phe Ile Pro
20 25 30

Val Phe Cys Lys Ala Met His Val Ala Gln Pro Ala Val Val Leu Ala
35 40 45

Ser Ser Arg Gly Ile Ala Ser Phe Val Cys Glu Tyr Ala Ser Pro Gly
50 55 60

Lys Ala Thr Glu Val Arg Val Thr Val Leu Arg Gln Ala Asp Ser Gln
65 70 75 80

Val Thr Glu Val Cys Ala Ala Thr Tyr Met Met Gly Asn Glu Leu Thr
85 90 95

Phe Leu Asp Asp Ser Ile Cys Thr Gly Thr Ser Ser Gly Asn Gln Val
100 105 110

Asn Leu Thr Ile Gln Gly Leu Arg Ala Met Asp Thr Gly Leu Tyr Ile
115 120 125

Cys Lys Val Glu Leu Met Tyr Pro Pro Pro Tyr Tyr Leu Gly Ile Gly
130 135 140

Asn Gly Thr Gln Ile Tyr Val Ile Asp Pro Glu Pro Cys Pro Asp Ser
145 150 155 160

Asp Phe Leu Leu Trp Ile Leu Ala Ala Val Ser Ser Gly Leu Phe Phe
165 170 175

Tyr Ser Phe Leu Leu Thr Ala Val Ser Leu Ser Lys Met Leu Lys Lys
180 185 190

Arg Ser Pro Leu Thr Thr Gly Val Tyr Val Lys Met Pro Pro Thr Glu
195 200 205

Pro Glu Cys Glu Lys Gln Phe Gln Pro Tyr Phe Ile Pro Ile Asn
210 215 220

<210> 4
<211> 15
<212> DNA
<213> Homo sapien

<400> 4
agatcctyaa agtga 15

<210> 5
<211> 15
<212> DNA
<213> Homo sapien

<400> 5
cagtcaargg cagtg 15

<210> 6
<211> 15
<212> DNA
<213> Homo sapien

<400> 6
cactgagytg acacc 15

<210> 7
<211> 15
<212> DNA
<213> Homo sapien

<400> 7
ctagaacygt aggca 15

<211> 15
<212> DNA
<213> Homo sapien

<400> 13
ataaatcact gccyt 15

<210> 14
<211> 15
<212> DNA
<213> Homo sapien

<400> 14
ccatttcact gagyt 15

<210> 15
<211> 15
<212> DNA
<213> Homo sapien

<400> 15
gcaacaggtg tcarc 15

<210> 16
<211> 15
<212> DNA
<213> Homo sapien

<400> 16
aacgcactag aacyg 15

<210> 17
<211> 15
<212> DNA
<213> Homo sapien

<400> 17
tgccaatgcc tacrg 15

<210> 18
<211> 15
<212> DNA
<213> Homo sapien

<400> 18
aataaatttt aatrg 15

<210> 19
<211> 15
<212> DNA
<213> Homo sapien

<400> 19
ttcttgattc agcya 15

<210> 20
<211> 15
<212> DNA
<213> Homo sapien

<400> 20
ctgtatgctg tgarc 15

<210> 21
<211> 15
<212> DNA
<213> Homo sapien

<400> 21
ttaaaaatga atgyt 15

<210> 22
<211> 10
<212> DNA
<213> Homo sapien

<400> 22
tccagatcct 10

<210> 23
<211> 10
<212> DNA
<213> Homo sapien

<400> 23
tgttcacttt 10

<210> 24
<211> 10
<212> DNA
<213> Homo sapien

<400> 24
cagcagtcaa 10

<210> 25
<211> 10
<212> DNA
<213> Homo sapien

<400> 25
aatcactgcc 10

<210> 26
<211> 10
<212> DNA
<213> Homo sapien

<400> 26
tttcactgag 10

<210> 27
<211> 10
<212> DNA
<213> Homo sapien

<400> 27
acaggtgtca 10

<210> 28
<211> 10
<212> DNA
<213> Homo sapien

<400> 28
gcactagaac 10

<210> 29

<211> 10
<212> DNA
<213> Homo sapien

<400> 29
caatgcctac 10

<210> 30
<211> 10
<212> DNA
<213> Homo sapien

<400> 30
aaattttaat 10

<210> 31
<211> 10
<212> DNA
<213> Homo sapien

<400> 31
ttgattcagc 10

<210> 32
<211> 10
<212> DNA
<213> Homo sapien

<400> 32
tatgctgtga 10

<210> 33
<211> 10
<212> DNA
<213> Homo sapien

<400> 33
aaaatgaatg 10

<210> 34
<211> 18
<212> DNA
<213> Homo sapien

<400> 34
tgtaaaacga cggccagt 18

<210> 35
<211> 19
<212> DNA
<213> Homo sapien

<400> 35
aggaaacagc tatgaccat 19

<210> 36
<211> 9336
<212> DNA
<213> Homo sapien

<220>
<221> allele
<222> (3735)
<223> PS1: polymorphic base C or T

<220>
<221> allele
<222> (4102)
<223> PS2: polymorphic base A or G

<220>
<221> allele
<222> (4238)
<223> PS3: poolymorphic base A or G

<220>
<221> allele
<222> (7067)
<223> PS4: polymorphic base T or C

<220>
<221> allele
<222> (7747)
<223> PS5: polymorphic base C or T

<220>
<221> allele
<222> (8660)
<223> PS6: polymorphic base A r G

<220>
<221> allele
<222> (8708)
<223> PS7: polymorphic base A or G

<400> 36
actcaaattt cccccagttt ctcagtaatc tgctttgtat tatttttttc aatcaataat 60
aacacattaa attctgtcat catgtctcta ttttctcctt taatctagaa tggttttcca 120
ggctttttcc tttttaatct tttgtgatgt taacattttt gaagaatcta ggccagttat 180
tttgcagaat atggatttgt ctgattattt cttcatgctt agattcagtt aaaacaattt 240
tggcaataac tacacaaagt tgttgtgtcc ttctcagtgt attgtatctg gaggtaccct 300
atgtcagaac actccatcac agtaatcttc agttgggatc acttggttaa agcagtatct 360
gctagattta tcaattatct tttccctatt gtaattatca agtaatctgt ggggtggtat 420
atgtgaatat atctttttct accaagattc ttttacccag catttttttt acaaccattt 480
gaattttctt gctatgacag aattagtttc ataagagcat aattaaggaa tctggtaaga 540
cgtctttgaa agttctaatt acaaccctca tctcttctta tgttaaactg tcttcattca 600
tttacctacc tatttaacaa atgtttaaag aatgctcgtt ttgtggcttg gcctgtgaca 660
ttttcatcag tggcttcata gttgcctaaa ctaatgtctc ttcctgtgct ccagaatcct 720
attggcaact attttccatg gatgccataa actcaacata ctcagatgaa agctgtcatt 780
tcactttctc acctgctctt cctgttattt attccagtaa attgcacctc tgtcttctca 840
ttgaccctat aataagaaag gcctatttat tgagtggcta tctgttacaa taggcctttt 900
atacatgtta catgcattat ctcatttaat cttttcaata atcttgagaa ttagatgcca 960
ttatctggag gacattgggc tttgagcccc aggtcaacag taagcaagtc cccatgctag 1020
gatacaaatc cagttatctt gggcttcaaa atctttgctt tagcatagtg ataaaggtat 1080
cactggcttc aaatcctggc tctgccactt cctacgggtg attgggtaag tcaccttaac 1140
tttcaatgcc ttgatttcct tctttataaa atgggaaaaa tggtaactct tgtcttgtag 1200
ggttgttatg gacttgaaag aaacaaaact tgctgggcgc ggtggctcac gcctgtaatc 1260
ccagcgcttt gggaggccga ggcgggtaga tcacctgagg tcaggaattt gagactagcc 1320
tggtcaacat ggcgaaaccc catctctact aaaataatac aaaaaattag ctgggcgtgg 1380
tggtgggtgc ctgcaatccc agctacttgg gaggctgagg caggagaatc acttgaaccc 1440
aggaggcaga ggttgcagtg agccgagatc gcgccactgc actccaacct gggaaacaag 1500
agtgaaactc tgtctcaaaa aaataaaaaa gaaaagaaag aaaacttgac atttgaataa 1560
tttctgcagc accacagtac ttctcaaacc ataattctgg tgtcattcat tctaaacttc 1620
tccttccact cactgattgt ggactagtcc catagggacc tcgtcttcac agctgtcatt 1680
accttagttt agaccagggg ttggcaaacc atggcccaca ggccaaatcc agacttcacc 1740
tattttcgta aatgaagttt tattggaatg cacacacacc catttgttta catataatat 1800
atggcttctt tcattctaaa atgacagagt tgagtagtgg caacagagac cccaccgttt 1860
gcaaatcata acatatttac tattttgccc ccttcagaaa gctttccaaa tcctggtgcc 1920
tggcgtattc caatagtctt cttcccagtc ttcctggtta catttctccc tgaacccatc 1980
ctccccatcc ctaaaatttc ccccaaatgt gaactcagtc ataccagttt gctccttatg 2040
tttcattggc ccttgctgct aagagcatcc gcttgcacct tctgctcatc cccagacaag 2100
ctttgtcctg tgaccataat gaactcttca tgccgtttcc aactttagcc catgttattc 2160
ttcttgtctg aatatccacc cttttctctg ttctcaataa taagttcagg cttttcgtct 2220
tctgagaagc cctttctgac ttccacaggc tgaaccactg gcttctgctc ctctacataa 2280
tacttcaatt ccagcattga tctcactcta tcatgatcat gggtttagct gtctgtccct 2340
gccactgctg tgtgttcctc ttgagggcag gaacatttgt ttttcacttt ttaaaaaacc 2400
tctgttgccc agtctggcat taggaagtgc ccattaggtt gttattgctt gttggcgctt 2460

gagctggggc ttgaaggttt ctataatgtg tagcagtgta tagaaaacag gcaggtcaga 2520
aaaggcttct gtgcatcaca ccaacatggc acatgtatac atatgtaaca aatctgcatg 2580
ttgtgcacat gtaccctaaa acttaaagta taataataat aaaattttaa aaaaaaaaag 2640
aagaggcttc ctggaggaga tgacagctga gctaagtcct ggaggatgag aaggagtata 2700
aaataagata ataggagaaa aaaggcagta ggaacagcat gggtaaaggt gatgaggcct 2760
gaaagaggca cgtggaagga aagacaaatg caggaagggg gaatgggagg gaatgctggg 2820
gtacaggcca aagagggagg cattcgggga gtgttctttt aaacatcctt tggtgtgcta 2880
atgtgtggtc attgggaaac catggaccgt tttttttgg caaacctcat atcccctgtt 2940
acaactgtct gtttgcatgt cagccttgta aaagcccctt aaggtatcaa ctatgttttt 3000
gttttgtcat cattcaatcc taagtgcaca gaattccggg catattacag gttccccatg 3060
aatgtttctt tctttattaa aatgtatgaa aactctccag atttaaggaa ggtcctcaat 3120
gtttcaaatt ctttttgtta gatcattggt cctgtctaca gctgtcacaa atttaaggac 3180
tctggttata tttaatcttc acttttgaat tttctgcttg aaaaatttgt attagaaaaa 3240
aaagtctatc cttttatgga cggctctaat ctcttgaatc atttgggttg gcttttcttt 3300
ggaccttctt caactctgtt ttgtctctgt tgagttaagg cttttaagaa cacctgaatt 3360
ctttccttct gcaaaaccag aggcagcttc ttttccgcct attttcagtt tatttcttgt 3420
gattttagtt tttttctctt aaccaaatgc taaatggatt taggagaaat aaacttattt 3480
gtaaagctgt caagggacca ttagaaggat ggtgcttcac agatagaata cagtttttat 3540
taatgatgcc tagacaaatc ctgccattag cccaagggct cagaaagtta gcagcctagt 3600
agttttggag ttgtcaatga aatgaattgg actggatggt taaggatgcc cagaagattg 3660
aataaaattg ggatttagga ggacccttgt actccaggaa attctccaag tctccactta 3720
gttatccaga tcctyaaagt gaacatgaag cttcagtttc aaattgaata cattttccat 3780
ccatggattg gcttgttttg ttcagttgag tgcttgaggt tgtcttttcg acgtaacagc 3840
taaacccacg gcttcctttc tcgtaaaacc aaaacaaaaa ggctttctat tcaagtgcct 3900
tctgtgtgtg cacatgtgta atacatatct gggatcaaag ctatctatat aaagtccttg 3960
attctgtgtg ggttcaaaca catttcaaag cttcaggatc ctgaaaggtt ttgctctact 4020
tcctgaagac ctgaacaccg ctcccataaa gccatggctt gccttggatt tcagcggcac 4080
aaggctcagc tgaacctggc trccaggacc tggccctgca ctctcctgtt ttttcttctc 4140
ttcatccctg tcttctgcaa aggtgagtga gacttttgga gcatgaagat ggaggaggtg 4200
tttctcctac ctgggtttca tttgtttcag cagtcaargg cagtgattta tagcaaagcc 4260
agaagttaaa ggtaaaactc caatctggct tggctggctc tgtattccag ggccagcagg 4320
gagcagttgg gcggcacgaa ataaggcaaa gagatagctc agaacagagc gccaggtatt 4380
tagtaggggc ttcatgaatg catgtgagtt ggtttagtag agagacacag gcaatttcag 4440
acccttctat gagactggaa gtgatttaag agggaaagga tagccatagt cctgaataca 4500
tttgagctgg gtttcaggat gagctcacaa gttcctttaa aaaaaattga cttaagcaaa 4560
tcctgggaag agttttttttg ctatacaatt caaggtttta aggtcctcgg attcatatac 4620
tttataaatg aattagccag cttgtttaaa atgtagggaa attgtgggaa gaatgccttc 4680
tttacttaat tcaaggtttt aaggttctct taatcaattc tactagctaa ttagccaatt 4740
atttaaaaat aaaagtttga aattgccaaa aaaaaaagac aaggaaaagg aaagaaagaa 4800
agccaccagt ctgtttggca tacaatactt aattgttgcc tgacctacgt gtgggtttca 4860
gatgcagatc ctcagttttc agctcttcag agactgacac caggtttgtt acacggctta 4920
aaatgatgag tatatccatt gaatctcaac cttatctctc tctagacctt cttggttaag 4980
aaaccatgta gtttgtatga agtaggtact caaaagatat ttgatgattt aatttttact 5040
ggagaagaaa tattcatata tgttttctta tttttacatg ttttaaatat gtaaagatta 5100
aataaacact cttagaagta tttaaatttc ctaaagtaaa tttatctcaa ccagtaacag 5160
gaccctccca atactggaaa gttgagtgtg accgcattta gtggtgatga gtgtgagctt 5220
gcttggggag agggcaggac atttaggatt tcttaagctt agagtcaata caataaagat 5280
tattgagtgc tcacttgggt gggctataat cactgctcac aggagttcat gaaccacaag 5340

taaaagagtg aggagatatg attagctcac aaataacttt aatacagagc agaaagtaat 5400
gaactactgc aatggagtta tcacagtgct aaggatgctc agagggcatc tctgataggc 5460
agaggtgagg gttagggaag gaagctgtag tctagctagc tagagctgct ggaatagaca 5520
tgacaatggc tgctgcaaac tgttttctct tctgaggaca gatgtcccgt gcaagtggct 5580
tggtggaagg gactagtgtc tctaatatag ggtgatttat aagcaggaaa gtgtgtccta 5640
gaaattcaga ccagagtgat agattggaat tggatcatgg gggactcatt gaatgttatt 5700
tattgtattt gtttttgcga tcagtgttag taaagtgtca aagggattga gcagatgagt 5760
gacatcatgc aacacaagtt ttgagtttca cttgtcagac tgactggaga ggggcctggt 5820
tagttacagg aaggtaattt ggcatgcagc cactattttt gagttgatgc aagcctctct 5880
gtatggagag ctggtctcct ttatcctgtg ggaaaagaga acaaaggagc atgggagtgt 5940
tcaagggaag gagaaataaa gggcagagag gcagcggtgg tgtcagggga agcccacagg 6000
agttaacagc agggttgcct caacctagag aggaagcgac ctggtgccct cggctctgtg 6060
gcttccttca tctaacaaca tcttccactc tacaacaatg ccagggaagg cggaggctgg 6120
tacagtgcat caagacacag ctactcctgg gtgacagagg ttcagggcca gctcactaag 6180
taggcagaag tttttgacat atactttgag agataaagca agattctgta cctcaacctt 6240
cagaatttcc cctaccactc attatagttc cggagctata tagctcctat cattctatca 6300
taaccttaga ataccagaga acatatcatc tcatctaatt atctcttact atatgtgaaa 6360
aaaatgaagg acatggggga agtgtgactt gccccaaatc acatatttca tggtagagcc 6420
aggtcttctg tttgtcatat cagtgttctt cctgccacaa ccatcttgaa gaatctattt 6480
ctcagtaaga aaatatcttt atggagagta gctggaaaac agttgagaga tggaggggag 6540
gctgggggtg tggagagggg aaggggtaag tgatagattc gttgaagggg ggagaaaagg 6600
ccgtggggat gaagctagaa ggcagaaggg cttgcctggg cttggccatg aaggagcatg 6660
agttcactga gttccctttg gcttttccat gctagcaatg cacgtggccc agcctgctgt 6720
ggtactggcc agcagccgag gcatcgccag ctttgtgtgt gagtatgcat ctccaggcaa 6780
agccactgag gtccgggtga cagtgcttcg gcaggctgac agccaggtga ctgaagtctg 6840
tgcggcaacc tacatgatgg ggaatgagtt gaccttccta gatgattcca tctgcacggg 6900
cacctccagt ggaaatcaag tgaacctcac tatccaagga ctgagggcca tggacacggg 6960
actctacatc tgcaaggtgg agctcatgta cccaccgcca tactacctgg gcataggcaa 7020
cggaacccag atttatgtaa ttggtgagca aagccatttc actgagytga cacctgttgc 7080
attgcagtct tctatgcaca aaaacagttt tgttccttaa tttcaggagg tttacttta 7140
ggactgtgga cattctcttt aagagttctg taccacatgg tagccttgct tattgtgggt 7200
ggcaacctta atagcattct gactgtaaaa taaaatgatt tggggaagtt ggggctctcg 7260
ctctggagtg ctaaccatca tgacgtttga tctgtacttt tgatatgata tgatgctcct 7320
ggggaagtag tcccaaatag ccaaacctat tggtgggcta cccatgcaat ttaggggtgg 7380
acctcaaggc ctggaagctc taatgtcctt ttttcaccaa tgttggggag tagagcccta 7440
gagtttaaaa ctgtctcagg gaggctctgc tttgttttct gttgcagatc cagaaccgtg 7500
cccagattct gacttcctcc tctggatcct tgcagcagtt agttcggggt tgtttttta 7560
tagctttctc ctcacagctg tttctttgag caaaatggtg agtgtggtgc tgatggtgca 7620
ccatgtctga tggggatacc tttagtggta tcaactggcc aaaagatgat gttgagttta 7680
gtgttcttga gatgagatga ggcaataaat gaagaggaag gacagtggta aagaacgcac 7740
tagaacygta ggcattggca tttgaggttt cagaatgact aatattttag atgaatttgt 7800
ttgacattga atgttcatgt gcttctgagc agggtttcaa tttgagtaac cgttgcaata 7860
acatggggca gctgttttgc tctttgtctt catgacaact gtacttaagc taacagccct 7920
gaaacatgag attaggctgg gcagaatgct gctagagagg accacttgga tggtctttat 7980
tctccttctc catgtccctc tccatcacct ggaagtcacc tctgggtgcc actctggtgc 8040
cttccttgtc gaagctgtag ctgctcacat gacacctatc cctgttatcc agtttgcttg 8100
actgggacgt tttgccttcc ccttcagcca ggaagtgaaa gtcccagttt ttatttatca 8160
caggtgttgg tattggtggt agaagaggta gaattatgga atcaggcctc ctgtcaggat 8220